UNIVERSIDADE DE LISBOA

FACULDADE DE MEDICINA

Quality control of gene expression in the mammalian cell nucleus

Noélia Maria Fernandes Custódio

Doutoramento em Ciências Biomédicas
Especialidade de Ciências Morfológicas

2008

A impressão desta dissertação foi aprovada pela Comissão Coordenadora do Conselho Científico da Faculdade de Medicina de Lisboa em reunião de 16 de Outubro de 2007

UNIVERSIDADE DE LISBOA
FACULDADE DE MEDICINA

Quality control of gene expression in the mammalian cell nucleus

Noélia Maria Fernandes Custódio

Tese orientada pela Professora Doutora Maria Carmo-Fonseca

Doutoramento em Ciências Biomédicas
Especialidade de Ciências Morfológicas

2008

As opiniões expressas nesta publicação são da exclusiva responsabilidade do seu autor

Noélia Maria Fernandes Custódio

Controlo de qualidade da expressão génica no núcleo de células de mamífero

Dissertação de Candidatura ao Grau de Doutor em Ciências Biomédicas, especialidade de Ciências Morfológicas, apresentada à Faculdade de Medicina da Universidade de Lisboa

Lisboa, 2008

Este trabalho foi parcialmente financiado pela Fundação para a Ciência e a Tecnologia, no âmbito do Programa PRAXIS XXI, através da Bolsa de Doutoramento PRAXIS XXI/BD/9439/96

TABLE OF CONTENTS

PREFÁCIO

Nesta dissertação apresentam-se os resultados do trabalho de investigação desenvolvido entre os anos de 1997 e 2007 na Faculdade de Medicina da Universidade de Lisboa, sob orientação da Professora Doutora Maria Carmo-Fonseca. O trabalho foi iniciado no Instituto de Histologia e Embriologia da Faculdade e continuado, desde 2003, na Unidade de Biologia Celular do Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa.

Este trabalho teve como principal objectivo elucidar o mecanismo de controlo de qualidade que leva à retenção nuclear dos transcritos do gene da β-globina humana com mutações que afectam o processamento do pré-mRNA.

Como previsto no Artigo 40º do Regulamento de Estudos Pós-graduados da Universidade de Lisboa (Deliberação nº 961/2003), a presente dissertação encontra-se redigida em língua inglesa, contendo um resumo alargado (mais de 1200 palavras) em língua portuguesa.

A dissertação encontra-se dividida em três capítulos: no primeiro capítulo - *Introduction*, é feita uma revisão geral sobre as vária etapas da expressão génica com principal incidência nas etapas que ocorrem ao nível do núcleo. Inicialmente é feita uma descrição de cada uma das etapas e da maquinaria nela interveniente. É dada especial relevância às inter-relações que hoje se conhecem entre as várias etapas da expressão génica e é feita uma revisão sobre o seu controlo de qualidade ao nível do núcleo. Finalmente apresenta-se o modelo utilizado e os principais objectivos deste estudo.

No segundo capítulo - *Results*, são apresentados os resultados originais obtidos neste estudo sob a forma de artigos publicados.

Finalmente, no terceiro capítulo - *Concluding Remarks and Future Pespectives*, são realçadas as principais conclusões do trabalho e expostas as perspectivas que se abrem para futuros estudos.

Como previsto no Decreto de Lei 388/70, art. 8º, parágrafo 2, parte integral dos resultados encontra-se publicada nos seguintes artigos:

Custódio, N., Carmo-Fonseca, M., Geraghtly, F., Pereira. S.H., Grosveld, F., Antoniou, M. (1999) Inefficient processing impairs release of RNA from the site of transcription. *EMBO J*. **18**: 2855-2866.

Custódio, N., Carvalho, C. Condado, I., Antoniou, M., Blencowe, B., Carmo-Fonseca, M. (2004) *In vivo* recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei. *RNA* **10**: 622-633.

Custódio N, Antoniou M, Carmo-Fonseca M. (2006) Abundance of the largest subunit of RNA polymerase II in the nucleus is regulated by nucleo-cytoplasmic shuttling. *Exp Cell Res*. **312**: 2557-2567.

Custódio, N., Vivo, M., Antoniou, M. and Carmo-Fonseca, M. (2007) Splicing- and cleavage-independent requirement of RNA polymerase II CTD for mRNA release from the transcription site. *J. Cell Biol.* **179**: 199-207.

A realização deste trabalho não teria sido possível sem a colaboração de pessoas e instituições a quem desejo deixar expressos os meus agradecimentos.

Em primeiro lugar, gostaria de endereçar um agradecimento especial à Professora Doutora Maria Carmo-Fonseca, pelo privilégio que me concedeu ao admitir-me como membro da sua equipa de investigação e mais tarde como membro da equipa de docentes de Biologia Molecular da Célula. Agradeço-lhe ainda pelas excelentes condições de trabalho que me proporcionou e pelo empenho com que sempre orientou o meu trabalho e a minha formação científica.

Ao Professor Doutor J.F. David-Ferreira agradeço o facto de me ter recebido enquanto Director do Instituto de Histologia e Embriologia da Faculdade de Medicina da Universidade de Lisboa, para a realização do meu estágio de licenciatura.

Ao Doutor Michael Antoniou agradeço a sua colaboração no trabalho apresentado nesta dissertação e o facto de me ter recebido no seu laboratório no *Department of*

Experimental Pathology, Guy's Hospital Campus, Londres, onde efectuei um estágio de alguns meses em 1998 e 99. Os seus conselhos, assim como os dos seus colaboradores Selina Raguz e Desmond Chow, foram imprescindíveis à realização de uma parte do trabalho aqui apresentado.

Os trabalhos publicados apresentados nesta dissertação contaram ainda com a participação dos seguintes co-autores: Helena Sofia Pereira, Inês Condado, Célia Carvalho, Maria Vivo (Instituto de Histologia e Embriologia e Instituto de Medicina Molecular), Finola Geraghty, Michael Antoniou (King's College London School of Medicine, London, UK), Frank Grosveld (Erasmus University, The Netherlands), Benjamin J. Blencowe (University of Toronto, Canada) a quem agradeço a valiosa contribuição.

Agradeço à Fundação para a Ciência e a Tecnologia o apoio financeiro que me concedeu no âmbito de uma Bolsa de Doutoramento (Bolsa PRAXIS XXI/BD/9439/96) durante a fase inicial do trabalho aqui apresentado.

A todos os meus colegas de trabalho do Instituto de Histologia e Embriologia e mais tarde do Instituto de Medicina Molecular, quero agradecer a amizade constante, os conselhos sempre úteis e a troca de ideias e conhecimentos que ajudou a ultrapassar muitos obstáculos ao longo dos anos. Não posso deixar de expressar aqui uma palavra especial para todos aqueles que conheci como colegas e que hoje considero também amigos: Ana Garcia, Ângelo Calado, Anita Gomes, Fátima Almeida, Hélia Neves, Isabel Alcobia, Maria Vivo, Marta Agostinho, Patrícia Calado, Peter Jordan, Sandra Caldeira, Sandra Martins, Sofia Pereira, Teresa Raquel, Zé Braga e Zé Rino. Por fim, gostaria de agradecer aos meus colegas com quem partilhei durante muitos anos a docência de Biologia Molecular da Célula e com quem muito aprendi: Célia Carvalho, Joana Desterro, João Ferreira, Margarida Gama-Carvalho e Teresa Carvalho.

Desejo também deixar expresso um agradecimento muito especial à minha família, pelo seu apoio e confiança.

RESUMO

Nos eucariontes, a transcrição dos genes pela RNA Polimerase II (RNA Pol II) origina moléculas de RNA mensageiro precursoras (pré-mRNA), sendo necessárias várias etapas de processamento até à formação do mRNA funcional que é transportado para o citoplasma, onde serve de molde para a síntese proteica. Essas etapas de processamento consistem em três processos distintos: na adição de uma trifosfoguanosina metilada na extremidade 5' trifosfato, estrutura conhecida como *cap*; na remoção de sequências não codificantes presentes no pré-mRNA (intrões) e junção das sequências codificantes flanqueadores (exões), num processo designado por *splicing*; e finalmente na formação da extremidade 3' do mRNA que consiste na clivagem e na adição à extremidade 3'OH resultante, de múltiplos nucleótidos de adenina, num processo designado por poliadenilação. Cada uma destas etapas de processamento é levada a cabo por uma maquinaria própria. No caso da adição do *cap* são necessárias três enzimas: uma RNA 5' trifosfatase, uma guanililtransferase e uma metiltransferase. O mecanismo de *splicing* envolve a ligação sequencial ao pré-mRNA de várias partículas ribonucleoproteicas nucleares pequenas (snRNP) que, com o auxílio de vários factores proteicos, formam uma estrutura complexa denominada spliceossoma. A maquinaria de poliadenilação é constituída por vários complexos proteicos, pela polimerase de poly(A) (PAP) e pela proteína nuclear de ligação ao poli(A) (PABPN1, anteriormente designada PABP2) (revisto por Proudfoot et al., 2002).

O correcto processamento das moléculas de pré-mRNA é um requisito fundamental para o transporte do mRNA para o citoplasma. Existem evidências de que moléculas de pré-mRNA com mutações nas regiões de consenso 5' e 3' dos seus intrões, que permitem a formação do spliceossoma mas que impedem a remoção dos intrões, são retidas no núcleo, enquanto que moléculas de pré-mRNA com mutações que não permitem a formação do spliceossoma são rapidamente exportadas para o citoplasma (Chang and Sharp, 1989; Hamm and Mattaj, 1990; Legrain and Rosbash, 1989). Do mesmo modo, mutações que interferem com a correcta poliadenilação do pré-mRNA também impedem a sua exportação para o citoplasma (Antoniou et al., 1998). Apesar destes resultados apontarem para a existência, no núcleo, de um mecanismo de controlo da qualidade do mRNA exportado, sabe-se muito pouco sobre o modo como esse controlo é efectuado. Com o objectivo de compreender o mecanismo de controlo de qualidade do mRNA exportado em células de mamífero iniciou-se

o estudo da organização intranuclear do RNA da β-globina humana. O gene da β-globina humana foi escolhido para este estudo devido ao grande número de mutações pontuais que afectam o processamento do pré-mRNA e que estão na base de uma doença genética designada por β-talassémia (Antonioiu, 1995). A expressão deste gene está limitada a células da linhagem eritróide, durante o processo de diferenciação (Tang et al., 2002). Assim, como modelo de estudo, utilizaram-se células de eritroleucemia murina (MEL, Murine ErythroLeukemia), que podem ser induzidas a diferenciar-se em cultura e expressar os genes específicos das células eritróides, onde se incluem os genes das globinas. Estas células foram transfectadas de forma estável com o gene da β-globina humana (βWT) e com mutantes do referido gene, nomeadamente com mutações que impedem o *splicing* do segundo intrão (βSM) e uma mutação que impede a poliadenilação (βIVSI). Quando este estudo foi iniciado sabia-se que os genes mutados eram transcritos e que o respectivo RNA não se acumulavam no citoplasma das células MEL, ao contrário do que acontecia com o mRNA da β-globina humana normal (Antoniou et al., 1998). O primeiro objectivo deste estudo foi então determinar a localização nuclear dos transcritos mutados. Para alcançar este objectivo foi optimizada a técnica de hibridação *in situ* para detecção dos transcritos do gene da β-globina humana nas células MEL. Esta experiência mostrou que os transcritos do gene normal eram detectados no núcleo, junto ao local de transcrição e acumulavam-se no citoplasma, enquanto que os transcritos dos genes mutados eram detectados apenas junto ao local de transcrição. Para avaliar a cinética de libertação do RNA do local de transcrição foram efectuadas experiências com o inibidor de transcrição actinomicina D. Este inibidor actua muito rapidamente após a sua adição ao meio de cultura (Darnell et al., 1971) e exerce o seu efeito através do bloqueio da elongação, uma vez que se intercala nas moléculas de DNA (Reich and Goldberg, 1964). Após 5 minutos de tratamento com actinomicina D os transcritos do gene da β-globina humana normal deixaram de ser detectados no núcleo, o que sugeriu que o RNA previamente transcrito já tinha sido libertado do local de transcrição. Por outro lado, os transcritos que têm uma mutação na região de consenso 5' do segundo intrão (βSM), que permite a formação do spliceossoma, mas inibe o *splicing* (Lamond et al., 1987) permaneceram junto ao local de transcrição após o tratamento com actinomicina D. Os transcritos que não possuem o segundo intrão (βIVSI) e por isso não são poliadenilados (Collis et al., 1990) ficaram também retidos no núcleo, junto ao local de transcrição, após o tratamento com actinomicina D. Estes resultados mostraram que os transcritos do gene da β-globina humana são rapidamente libertados do local de transcrição, enquanto que os

transcritos que possuem mutações que impedem o *splicing* e a poliadenilação ficam retidos próximo do gene, sugerindo que o mecanismo de controlo de qualidade do mRNA exportado actua ao nível do local de transcrição (Custódio et al., 1999).

O objectivo seguinte foi determinar qual o mecanismo molecular responsável pelo controlo de qualidade do mRNA ao nível do local de transcrição. Estudos efectuados na última década têm revelado a existência de inúmeras interacções entre a maquinaria de transcrição, de processamento do pré-mRNA e de exportação do mRNA para o citoplasma (ver Bentley, 2005). Uma das hipóteses colocadas para explicar a retenção dos transcritos com mutações junto ao local de transcrição foi a de que neste caso poderia não haver recrutamento de proteínas essenciais para a libertação e/ou transporte do mRNA a partir do respectivo gene. A outra hipótese colocada foi a de que a retenção poderia ocorrer por intermédio de proteínas que estão ligadas aos transcritos nascentes durante o processamento e que se libertam quando este se completa, mas que no caso dos mutantes ficam bloqueadas devido ao facto de o processamento do pré-mRNA não se completar.

Para testar a primeira hipótese decidimos estudar o recrutamento, para o local de transcrição, de proteínas que se ligam ao mRNA. Assim testamos o recrutamento de proteínas que se ligam ao mRNA após o *splicing,* designadas de proteínas do complexo junção de *splicing* (EJC, *exon junction complex*), e de factores de exportação do mRNA. Para tal, conjugou-se a técnica de hibridação *in situ* para detecção dos transcritos com a localização das proteínas em questão (por imunomarcação com anticorpos específicos ou transfecção para expressão de proteína com marcador fluorescente). Assim, efectuou-se detecção de proteínas do EJC (REF/Aly, Y14, SRm160, UAP56, RNPS1 e Magoh), de componentes do spliceossoma (proteínas Sm e proteína B'' do snRNP U2) e dos factores de exportação do mRNA (NXF1/TAP e p15) simultaneamente com a detecção dos transcritos da β-globina humana. Os resultados obtidos indicaram que as proteínas EJC e os componentes do spliceossoma são recrutados para os locais de transcrição do gene normal, no entanto, não se detectou acumulação dos factores de exportação junto ao local de transcrição. Estes resultados sugerem que as proteínas do EJC se ligam estavelmente ao pré-mRNA co-transcricionalmete, mas que os factores de exportação se ligam imediatamente antes da libertação dos transcritos do gene ou após estes se terem libertado. Quando se efectuou a mesma análise num mutante de processamento não se detectou recrutamento para o local de transcrição, nem das proteínas do EJC, nem dos componentes do spliceossoma. Estes resultados sugerem que o *splicing* dos transcritos da β-globina é essencial para a acumulação das proteínas do EJC no local de

transcrição e possivelmente para o posterior direccionamento para a via de exportação do mRNA para o citoplasma (Custódio et al., 2004). Embora os resultados obtidos não descartem a hipótese de que o recrutamento eficiente de proteínas do EJC e/ou de factores de exportação do mRNA possa contribuir para a libertação do mRNA do local de transcrição, são ainda necessários mais estudos para que se possa concluir se estas proteínas desempenham um papel essencial neste mecanismo.

Sabe-se que a associação entre a maquinaria de transcrição e a maquinaria de processamento do pré-mRNA é mediada pela extremidade carboxílica da subunidade grande da RNA polimerase II (RNA Pol II LS), normalmente designada de CTD (Carboxyl-Terminal Domain) (revisto por Bentley, 1999). O CTD da RNA Pol II apresenta uma constituição invulgar, sendo formado por repetições em tandem de um heptapéptido com o consenso Tirosina-Serina-Prolina-Treonina-Serina-Prolina-Serina, seguido de um motivo C-terminal constituído por 10 aminoácidos (Corden et al., 1985). O número de heptapéptidos presentes varia entre os eucariontes, apresentando os mamíferos 52 unidades repetitivas (Corden et al., 1985) e as leveduras apenas 26 (Allison et al., 1985). As primeiras evidências para uma ligação entre o CTD e o processamento do pré-mRNA vieram de estudos que apontavam para a existência de uma interacção entre determinados factores de *splicing* (Chabot et al., 1995; Kim et al., 1997; Mortillaro et al., 1996), de poliadenilação (McCracken et al., 1997b) e *capping* (McCracken et al., 1997a) com o CTD da RNA Pol II LS. Essas evidências foram fortalecidas quando se mostrou que uma forma da RNA Pol II LS com apenas 5 unidades repetitivas no CTD não era capaz de assegurar nem o *splicing* nem o processamento da extremidade 3' do pré-mRNA (McCracken et al., 1997b). Estudos mais recentes mostraram que as unidades repetitivas 1-15 e 1-25 são capazes de assegurar o *capping*, mas não o *splicing* ou a formação da extremidade 3' do pré-mRNA, enquanto que as unidades 27-52 em conjunto com os 10 aminoácidos do domínio C-terminal asseguram o *capping*, o *splicing* e o processamento da extremidade 3' (Fong and Bentley, 2001). Foi também mostrado que a alteração do motivo C-terminal reduz a eficiência do *splicing* e da clivagem da extremidade 3' do pré-mRNA (Fong et al., 2003) e impede a libertação do RNA do local de transcrição (Bird et al., 2005).

Tendo em conta estes dados, a hipótese de que a retenção junto ao local de transcrição pode ocorrer por intermédio de proteínas que se ligam aos transcritos nascentes e não se libertam devido ao bloqueio no processamento do pré-mRNA ganhou um candidato preferencial, o CTD da RNA Pol II. Assim, a retenção poderá ocorrer por intermédio da

maquinaria de processamento ineficaz ou inactiva que permanece associada ao CTD da RNA Pol II, pressupondo-se que a maquinaria de processamento se associa ao CTD durante a reacção e se desliga do CTD no final da mesma, permitindo a libertação do RNA para posterior transporte para o citoplasma. No caso dos mutantes, em que ocorre associação da maquinaria de processamento mas a reacção não se completa devido à mutação, o pré-mRNA poderá permanecer ligado ao CTD via essa mesma maquinaria, acabando por ser degradado. Para testar esta hipótese a abordagem experimental escolhida consistiu na construção de um modelo celular onde a transcrição dos genes é efectuada por uma RNA Pol II com delecção parcial do CTD. Esta abordagem é facilitada pela existência de formas da RNA Pol II LS resistentes ao inibidor de transcrição α-amanitina, o que permite anular a RNA Pol II endógena. Foram estabelecidas linhas celulares estavelmente transfectadas com o gene da RNA Pol II LS com 5 repetições no CTD (Δ5), com 31 repetições (Δ31) e com 52 repetições (controlo, *wild-type*). Foi efectuada uma primeira triagem a estas linhas para avaliar a sua capacidade de transcrever na presença de α-amanitina. Nenhuma das linhas Δ5 testadas apresentou capacidade de transcrever o gene da globina murina β-major na presença de α-amanitina. Este resultado foi inesperado, uma vez que este mutante já tinha sido utilizado em estudos anteriores, onde foi mostrado que apresentava capacidade de transcrever o gene da β-globina humana sob controlo do promotor SV40 (McCracken et al., 1997b). No entanto, estudos posteriores onde se analisou a capacidade transcricional da RNA Pol II Δ5 por *run-on* em mais de 500 genes, indicaram que a este mutante não é funcionalmente activo para a transcrição de genes endógenos (Meininghaus et al., 2000). Estes resultados impossibilitaram a utilização dos clones Δ5 nos estudos posteriores. Das linhas Δ31 e *wild-type* que apresentaram capacidade de transcrever na presença de α-amanitina foi seleccionada uma para transfecção com o gene da β-globina humana. Cada uma das linhas foi transfectada quer com a versão normal do gene (βWT), quer com uma versão que possui uma mutação de *splicing* (βSM) para a qual se mostrou anteriormente que há retenção do respectivo RNA junto ao local de transcrição. Os transcritos da β-globina humana foram detectados no núcleo das células transfectadas como um foco que corresponde ao RNA junto do local de transcrição. Quando se efectuou o tratamento com o inibidor de transcrição actinomicina D os resultados indicaram que uma grande proporção dos transcritos βWT feitos pela RNA Pol II Δ31 ficavam retidos no local de transcrição. Apesar deste resultado, observou-se recrutamento de proteínas do EJC, incluindo Aly/REF, Y14 e SRm160 para os locais de transcrição. Uma análise bioquímica dos transcritos βWT feitos pela RNA Pol II Δ31 indicou que sofriam

splicing e processamento na extremidade 3' correctamente. Estes resultados indicam que a RNA Pol II LS com apenas 31 unidades repetitivas é capaz de transcrever mas não permite a libertação eficiente do mRNA do local de transcrição. Isto sugere que a estrutura do CTD é importante para a libertação do mRNA do local de transcrição, possivelmente porque assegura o recrutamento de proteínas importantes para uma maturação final do mRNA após o *splicing* e o processamento da extremidade 3', que o transforma numa partícula ribonucleoproteíca competente para ser exportada. Propomos assim que as unidades repetitivas em falta no mutante Δ31 da RNA Pol II poderão ser essenciais para o recrutamento destas proteínas (Custódio et al., 2007).

A RNA Pol II eucariota é uma enzima complexa, composta por 12 sub-unidades distintas, que se encontra presente na célula em níveis relativamente baixos (revisto por Shilatifard et al., 2003). A transcrição do pré-mRNA pela RNA Pol II envolve ciclos de fosforilação e desfosforilação do CTD da sua maior subunidade. O CTD encontra-se hipofosforilado quando a RNA Pol II se liga ao complexo de pré-iniciação e assume uma forma hiperfosforilada quando o transcrito tem cerca de 25 bases (O'Brien et al., 1994). A primeira fosforilação ocorre na Serina que se encontra na posição 5 do heptapéptido e durante a elongação a Serina na posição 2 é fosforilada (Komarnitsky et al., 2000; Lu et al., 1991). Pelas razões já mencionadas, estabelecemos linhas celulares murinas que expressam uma forma da RNA Pol II LS resistente ao inibidor de transcrição α-amanitina. Foi efectuada uma primeira triagem a essas linhas celulares para seleccionar aquelas que apresentavam a capacidade de sobreviver e transcrever na presença da α-amanitina, assegurando que nessas linhas a RNA Pol II LS exógena é funcional. Com o objectivo de estudar a localização sub-celular da RNA Pol II LS seleccionaram-se duas linhas com níveis diferentes de expressão da subunidade exógena. Resultados de imunomarcação e *western-blotting* indicaram que a RNA Pol II LS sobre-expressa se encontra predominantemente hipofosforilada e se acumula sobretudo no citoplasma. A forma activa (fosforilada na Serina 2 do heptapéptido), pelo contrário, encontra-se exclusivamente no núcleo, em níveis constantes, independentemente do grau de sobre-expressão da proteína. A presença, no citoplasma, da RNA Pol II LS sobre-expressa foi descrita num outro estudo onde se procedeu à sobre-expressão de uma forma conjugada à proteína fluorescente EGFP (Sugaya et al., 2000). Este resultado pode ser explicado por uma importação ineficiente para o núcleo do excesso de proteína produzida no citoplasma, ou por o excesso de proteína que é importado para o núcleo ser posteriormente exportado para o citoplasma. Para investigar a possibilidade da RNA Pol II LS poder

deslocar-se do núcleo para o citoplasma comparámos a distribuição sub-celular das formas endógena e exógena sobre-expressa, na presença de leptomicina B (LMB), um inibidor de exportação nuclear dependente do transportador CRM1 (Fornerod et al., 1997; Kudo et al., 1999). Os resultados indicaram que o LMB diminui a acumulação citoplasmática da proteína sobre-expressa e aumenta os níveis nucleares da proteína endógena. Estes resultados indicaram, pela primeira vez, que a RNA Pol II LS tem a capacidade de efectuar um mecanismo de vai-e-vem entre o núcleo e o citoplasma. Propomos que este mecanismo poderá ser um meio a que a célula recorre para regular o número do moléculas desta sub-unidade presentes no núcleo, de modo a manter uma relação equilibrada das várias sub-unidades e assim assegurar a formação de complexos de transcrição funcionalmente activos (Custódio et al., 2006).

Em conclusão, o trabalho apresentado nesta dissertação foi pioneiro na identificação de um mecanismo de controlo de qualidade do mRNA, que actua ao nível do local de transcrição, no núcleo de células de mamífero (Custódio et al., 1999). Este mecanismo evita que moléculas de pré-mRNA com mutações que impedem o seu correcto processamento cheguem ao citoplasma, onde poderiam original proteínas defeituosas com consequências negativas quer para a célula quer para o organismo. O trabalho apresentado foi também o primeiro a mostrar a importância da libertação do mRNA do local de transcrição, actualmente considerada uma etapa importante da síntese do mRNA (ver Vasudevan and Peltz, 2003). Estudos efectuados em *S. Cerevisiae* vieram mostrar que um mecanismo de controlo de qualidade semelhante também funciona em levedura, onde transcritos com defeitos na poliadenilação ficam retidos no núcleo, ao nível do local de transcrição (Hilleren et al., 2001; Jensen et al., 2001b; Libri et al., 2002; Thomsen et al., 2003). O trabalho aqui apresentado contribuiu ainda para a compreensão do mecanismo responsável pela retenção dos transcritos com defeitos no processamento junto ao local de transcrição. Mostrámos que estes transcritos não são capazes de recrutar eficientemente proteínas do complexo junção de *splicing* (Custódio et al., 2004) e que o CTD da RNA Pol II desempenha um papel importante na etapa de libertação do mRNA do local de transcrição (Custódio et al., 2007).

PREFACE

This thesis presents the results of research that was carried out between 1997 and 2007 at the Faculty of Medicine, University of Lisbon, under the supervision of Professor Maria Carmo-Fonseca. The work was started at the Institute of Histology and Embryology of the Faculty and continued, since 2003, in the Cell Biology Unit of the Institute of Molecular Biology, Faculty of Medicine, University of Lisbon.

The main goal of this work was to elucidate the quality control mechanism responsible for the retention within the nucleus of human β-globin transcripts with mutations that impaired their processing.

This thesis is divided in three chapters: the first chapter - *Introduction* - presents a general overview of the steps of gene expression with a special focus on those that occur within the nucleus. A special focus is given to what is currently known on the coupling between the steps of gene expression and to the quality control mechanisms of gene expression that operate in the nucleus. In the end of this chapter, the model system used in these work and the general goals of the study are presented. In the second chapter - *Results* - the original results of this work are presented as published journal articles. Finally, the last chapter - *Concluding Remarks and Future Perspectives* - summarizes the most important conclusions and future perspectives that have emerged from this dissertation.

The work presented in this dissertation was possible due to the contribution of people and institutions to whom I would like to express my gratitude.

I am deeply grateful to Professor Maria Carmo-Fonseca for giving me the opportunity to be a member of her research group. I acknowledge her for having provided me excellent working conditions and for her supervision and encouragement throughout the work presented in this dissertation.

To Professor J.F. David-Ferreira I thank the opportunity to start my reseach work, still as an undergraduate student, at the Institute of Histology and Embryology, Faculty of Medicine.

I am grateful to Dr. Michael Antoniou, for his collaboration in the work presented here and for the generosity with which he received me in his laboratory in the Department of Experimental Pathology, Guy's Hospital Campus, London.

Special thanks to all the co-authors of the work presented in this dissertation for their valuable contribution: Helena Sofia Pereira, Inês Condado, Célia Carvalho, Maria Vivo (Instituto de Histologia e Embriologia and Instituto de Medicina Molecular), Finola Geraghty, Michael Antoniou (King's College London School of Medicine, London, UK), Frank Grosveld (Erasmus University, The Netherlands), Benjamin J. Blencowe (University of Toronto, Canada).

I wish also to acknowledge Fundação para a Ciência e a Tecnologia, for supporting me with the Fellowship PRAXIS XXI/BD/9439/96.

I would like also to express my gratitude to all my colleagues over the years for their permanent help, support and friendship. A special note of gratitude to those I initially met as colleagues that today I consider my friends: Ana Garcia, Ângelo Calado, Anita Gomes, Fátima Almeida, Hélia Neves, Isabel Alcobia, Maria Vivo, Marta Agostinho, Patrícia Calado, Peter Jordan, Sandra Caldeira, Sandra Martins, Sofia Pereira, Teresa Raquel, Zé Braga e Zé Rino. Finally, a special word to my colleagues with whom I shared for many years the experience of teaching and with whom I learned so much: Célia Carvalho, Joana Desterro, João Ferreira, Margarida Gama-Carvalho e Teresa Carvalho.

Finally, I am deeply grateful to my family for their unconditional support.

ABSTRACT

Protein-encoding genes are transcribed in the nucleus by RNA polymerase II as precursor RNAs that undergo extensive processing before being translocated to the cytoplasm for translation by the ribosomes. This spatial and temporal separation between RNA and protein synthesis offers an immense opportunity for regulation and quality control.

When this study was initiated it was known from biochemical studies that human β-globin (*HBB*) transcripts with mutations that impaired splicing or 3' end formation were unable to be exported to the cytoplasm, a phenotype identical to that seen in β-thalassemia patients harbouring similar mutations (Antoniou et al., 1998). This result was consistent with the retention in the nucleus of incorrectly processed transcripts. Based on this initial observation we hypothesised the existence of a quality control mechanism to retain the incorrectly processed transcripts in the nucleus and proposed to elucidate the mechanism responsible for the observed retention. The first goal of this work was to determine the intranuclear localisation of the retained transcripts. To address this we used as a model system murine erythroleukemia (MEL) cells stably transfected with either wild-type or processing mutant *HBB* genes. The experimental approach was based on the direct visualisation of both normal and defective *HBB* transcripts using fluorescence *in situ* hybridisation and confocal microscopy. Nuclear transcripts of both wild-type and mutant *HBB* genes were detected only as intranuclear foci co-localising with the template gene locus. To determine the kinetics of transcript release from the site of transcription the cells were treated with the transcriptional inhibitors actinomycin D, α-amanitin and DRB. These drugs induced the rapid disappearance of nuclear foci corresponding to wild-type *HBB* RNA. In contrast, pre-mRNA mutants defective in either splicing or 3' end formation and which fail to be transported to the cytoplasm were retained at the site of transcription. These results indicated that the quality control mechanism responsible for the nuclear retention of incorrectly processed transcripts operates at the site of transcription and suggest that splicing and 3' end processing are rate limiting for release of mRNA from the transcription site.

The next goal of this work was to determine the molecular players involved in the quality control mechanism that operates at the transcription site. Studies over the past years have indicated that there is extensive coupling between transcription, pre-mRNA processing and nuclear export of mRNA (reviewed by Bentley, 2005). One hypothesis to explain the

retention of the mutant transcripts could be the absence of recruitment of proteins essential for the release and/or transport of the transcripts from the gene locus. Alternatively, the retention could be mediated by proteins that are bound to the nascent transcripts and become stalled due to incomplete processing.

To test the first hypothesis we decided to study the recruitment of exon junction complex (EJC) proteins and mRNA export factors to the retention sites in the nucleus. To achieve this we visualised the distribution of EJC proteins and RNA export factors relative to the sites *HBB* transcription in the nucleus. Using *in situ* hybridisation and confocal microscopy, we observed accumulation of EJC proteins (REF/Aly, Y14, SRm160, UAP56, RNPS1 and Magoh) and core spliceosome components (U snRNPs) at sites of wild-type *HBB* transcription. This suggests that EJC proteins bind stably to pre-mRNA co-transcriptionally. No concentration of the export factors NXF1/TAP or p15 was detected on nascent transcripts, arguing that in mammalian cells these proteins bind the mRNA shortly before or after release from the sites of transcription. Contrasting with the results obtained in MEL cells expressing wild-type *HBB* transcripts, mutant pre-mRNA defective in splicing and 3' end processing do not co-localise with SRm160, REF, UAP56 or Sm proteins. These results suggest that the accumulation of EJC proteins at transcription sites requires efficient processing of the nascent pre-mRNA. Although the results obtained do not discard the hypothesis that efficient recruitment of EJC proteins and/or export factors may contribute to the release of the transcripts from the transcription site, further studies are required to conclusively show or discard the involvement of these proteins in transcription site release.

In the second hypothesis the best candidate to mediate the retention is the carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II (RNA Pol II LS). The mammalian CTD comprises 52 heptad repeats followed by a terminal 10 amino acid motif (Corden et al., 1985). Several reports in the last decade showed that both splicing and 3' end processing factors can associate with the CTD of RNA Pol II (reviewed by Bentley, 2005). Since the processing mutants analysed are able to assemble some processing machinery, the release of the transcripts could be blocked by the stalled or abnormal processing machinery associated with the CTD. In order to test this hypothesis we generated cell lines that express either the wild-type *HBB* gene or a mutant version defective in splicing and an α-amanitin resistant form of RNA Pol II LS with either 52 (wt), 31 (Δ31) or 5 (Δ5) repeats of the CTD. When these cells were treated with actinomycin D to stop transcription the results showed that a large proportion of the wild-type *HBB* transcripts made by RNA Pol II with only 31 CTD

repearts were retained at the site of transcription. Despite this result we observed recruitment of several EJC proteins, includind Aly/REF, Y14 and SRm160 to the transcription sites and RNase protection assays showed that the *HBB* transcripts produced by RNA Pol II LS Δ31 are correctly spliced and 3' end cleaved. These results show that the form of RNA Pol II LS with only 31 repeats of the CTD is competent in transcription and processing but fails to allow the efficient release of the transcripts from the transcription site. These results provide evidence that mRNA release from the transcription site requires the heptad repeat structure of the CTD and we propose that the missing heptads in the truncated CTD mutant are required for binding of proteins implicated in a final co-transcriptional maturation of spliced and 3' end cleaved mRNAs into export-competent ribonucleoprotein particles.

Eukaryotic RNA polymerase II is a complex enzyme composed of 12 distinct subunits that is present in cells in low abundance (reviewed by Shilatifard et al., 2003). Transcription of mRNA by RNA polymerase II involves a phosphorylation/ dephosphorylation cycle of the CTD of the enzyme's largest subunit. The hypophosphorylated form assembles into pre-initiation complexes and the CTD becomes first phosphorylated on Serine at position 5 of the heptad repeat and during elongation on Serine at position 2 (Komarnitsky et al., 2000; Lu et al., 1991). As mentioned before we have generated stable murine cell lines expressing an α-amanitin resistant form of RNA Pol II LS. The cell lines generated were screened for survival and transcription in the presence of α-amanitin ensuring the functionality of the exogenous subunit. To characterise these cell lines we studied the localisation of the exogenous subunit and observed that over-expressed RNA Pol II LS was predominantly hypophosphorylated, soluble and accumulated in the cytoplasm in a CRM1-dependent manner. Our results further showed that the transcriptionally active form of RNA Pol II LS containing phosphoserine in position 2 of the CTD repeats was restricted to the nucleus and its levels remained remarkably constant. These results suggest that the nuclear-cytoplasmic distribution of RNA Pol II LS may be regulated by shuttling and we propose that this may provide a mechanism to control the pool of RNA polymerase subunits that is accessible for assembly of a functional enzyme in the nucleus.

KEYWORDS

Transcription;

RNA polymerase II;

CTD;

Pre-mRNA processing;

mRNA export;

mRNA surveillance;

ABBREVIATIONS

A - adenosine
ADAR - adenosine deaminase that act on RNA
Br-UTP - bromouridine triphosphate
C - cytidine
CBC - cap binding complex
CDAR - cytidine deaminase acting on RNA
cDNA - complementary DNA
Ceg1 - RNA guanylyltransferase
Cet1 - RNA triphosphatase
CF - cleavage factors
CFIA - cleavage and polyadenylation factor IA
CFIB - cleavage and polyadenylation factor IB
ChIP - chromatin immunoprecipitation
CoTC - co-transcriptional cleavage
CPF - cleavage and polyadenylation factor
CPSF - cleavage-polyadenylation-specificity factor
CstF - cleavage stimulatory factor
CTD - carboxyl-terminal domain
CTE - constitutive transport element
DRB - 5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole
DSIF - DRB sensitivity-inducing factor
dsRNA - double-stranded RNA
EJC - exon junction complex
ESE - exonic splicing enhancer
GFP - green fluorescent protein
GMP - guanosine-5'-monophosphate
GTFs - general transcription factors
GTP - guanosine-5'-triphosphate
HBB - human β-globin gene
hnRNP - heterogeneous nuclear ribonucleoprotein
HS - DNase I hypersensitive site
I - inosine
IGC - interchromatin granule clusters
IIa - hypophosphorylated form of CTD
IIo - hyperphosphorylated form of CTD
LCR - locus control region
MEL - murine erythroleukemia
MEX - mRNA export
Mlp - myosin-like protein
mRNP - mensenger ribonucleoprotein particle
MTR - mRNA transport
NELF - negative elongation factor
NMD - nonsense-mediated mRNA decay
NPC - nuclear pore complexes
Nups - nucleoporins
NXF - nuclear export factor proteins
Pab1p - yeast poly(A) tail-binding protein
PABP - poly(A) tail-binding protein
PABPC - cytoplasmic poly(A)-binding protein
PABPN1 - nuclear poly(A)-binding protein
PAN - poly(A)-specific nuclease
PAP - poly(A) polymerase
PML - promyelocytic leukaemia
Poly(A)$^+$ - polyadenylated
Pre-mRNA - precursor messenger RNA
P-TEFb - positive transcription elongation factor b
Py - polypyrimidine
RAT - ribonucleic acid trafficking
REF - RNA export factor binding protein
RNA Pol II - RNA polymerase II
RNA Pol II LS - largest subunit of RNA polymerase II
RNAi - RNA interference
rRNA - ribosomal RNA
RRP6 - ribosomal RNA processing 6
SFC - splicing factor compartments
snoRNA - small nucleolar RNAs
snRNA - small nuclear RNA
snRNP - small nuclear ribonucleoprotein particle
SR - serine/arginine
TF - transcription factor
Tpr - translocated promoter region protein
TREX - transcription and export complex
U - uridine
U2AF - U2 snRNP auxiliary factor
βLCR - *HBB* micro-locus control region

CHAPTER I

INTRODUCTION

1. The cell nucleus

The nucleus is the hallmark of all eukaryotic cells. It houses the genome of the cell serving as the repository of genetic information and functioning as the control center of the cell. The nuclear envelope, a double membrane that can be regarded as a specialized extension of the endoplasmic reticulum (ER), separates the contents of the nucleus from the cytoplasm and provides the structural framework of the nucleus (see Cooper and Hausman, 2007). Many essential processes take place in the cell nucleus, including DNA replication, repair and recombination as well as the initial steps of gene expression (transcription and RNA processing). Only the final stage of gene expression (translation) takes place in the cytoplasm (Cooper and Hausman, 2007). Prokaryotes do not have a nucleus but carry out many of the processes that take place inside the nucleus in eukaryotic cells. So what functional advantages can the nucleus bring to a cell? By separating the genome from the cytoplasm, the nuclear envelope provided an opportunity for the evolution of new regulatory possibilities for gene expression that are unique to eukaryotes. For example, whereas prokaryotic mRNAs are translated while their transcription is still in progress, eukaryotic mRNAs undergo extensive posttranscriptional processing before being transported to the cytoplasm for translation (Cooper and Hausman, 2007). This spatial and temporal separation between transcription and protein synthesis offers new opportunities for regulation and opens a new window for the control of the quality of gene expression.

2. Gene expression as a multistep process

Gene expression can be defined as the process by which the genetic instructions of the DNA are transformed into functional proteins in a cell. Eukaryotic gene expression is a complex multistep process that requires several complex cellular machines. Each machine is responsible for a specific step in this process, which includes transcription, several precursor messenger RNA (pre-mRNA) processing steps (5' capping, splicing, 3' end processing and editing), export of the mature mRNA to the cytoplasm and translation into protein sequence (Figure 1). Identifying the protein components of each of this cellular machines and understanding how they work has been a major goal of molecular biologists for the last 25 years. For that purpose, each of the major steps outlined above has been carried out

independently *in vitro*. However, in the last decade numerous studies have provided evidence that they are not independent reactions *in vivo*. In contrast to a simple linear assembly line in which the transcript is made and passes from one processing reaction to the next, it is now clear that the processing reactions are co-transcriptional and interlinked in such a way that they influence one another's specifically and efficiently. The emerging picture now is that a complex and extensively coupled network exists to coordinate the activities of the several gene expression machines (reviewed in Hirose and Manley, 2000; Maniatis and Reed, 2002).

The goal of this section is to introduce the major steps in gene expression as well as the major players in each one of the steps. The connections between them will be the focus of section 3.

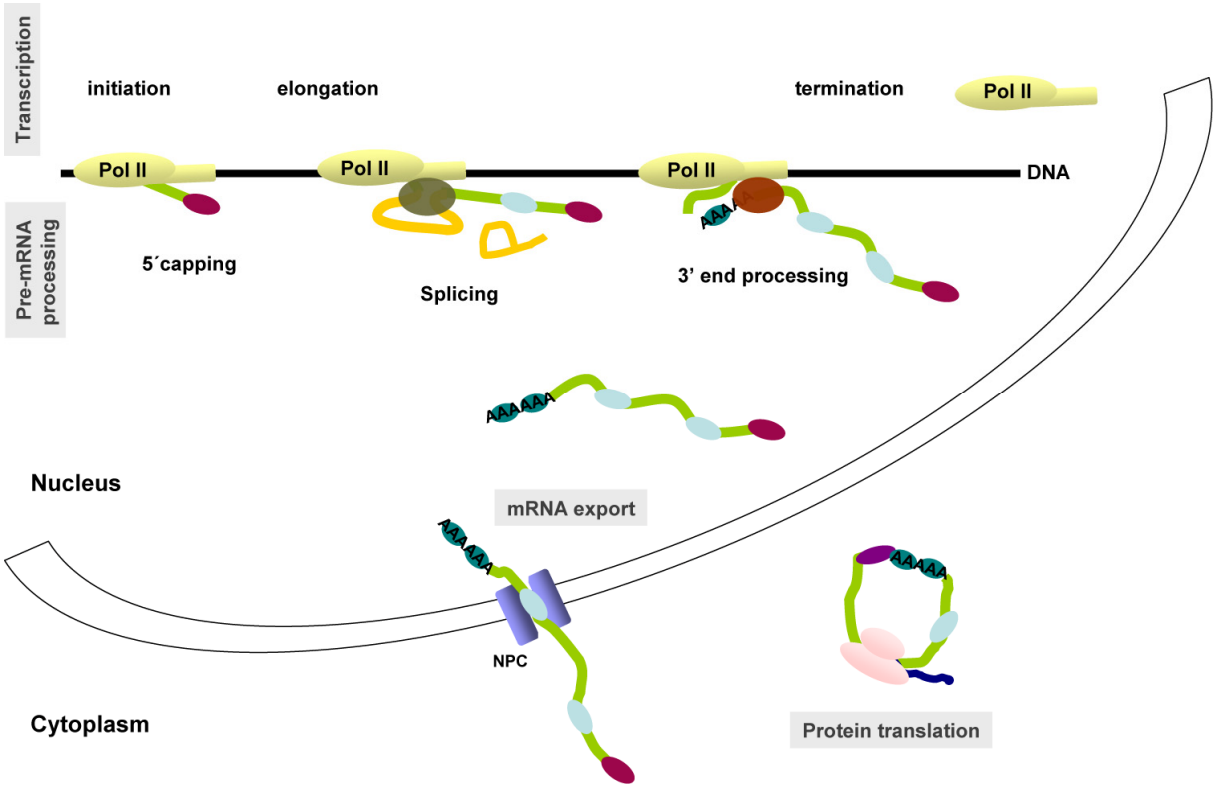


Figure 1 - The multiple steps of gene expression. The different steps in the pathway from gene to protein include transcription (initiation, elongation and termination), several precursor messenger RNA (pre-mRNA) processing steps (5' capping, splicing and 3' end processing), export of the mature mRNA to the cytoplasm and translation into protein sequence. In the figure the mRNA is represented by a green line and the introns by a yellow line. The cap binding complex (CBC) added to the 5' end of the mRNA is represented by a purple oval, the splicing machinery by a green oval and the 3' end processing machinery by a red oval. The blue ovals on the mRNA represent the several proteins that bind to it forming a messenger ribonucleoprotein particle (mRNP). The ribosome is represented by two pink ovals and the nascent peptide by a blue line (see text for more details).

2.1 Transcription by RNA polymerase II

Transcription, the DNA-directed synthesis of RNA, is the first step in the sequence of events that lead to gene expression. The many thousands of genes coding for proteins in eukaryotes are transcribed by the RNA polymerase II (RNA Pol II) transcription machinery. In a simplistic view, this machinery can be divided in three major components: the 12-subunit polymerase (Figure 2A), capable of synthesizing RNA and proofreading the nascent transcript; a set of five general transcription factors (GTFs), denoted TFIIB, -D, -E, -F and -H, which are responsible for promoter recognition and for unwinding the promoter DNA; and the Mediator complex, composed of 20 subunits, which transduces regulatory information from activator and repressor proteins to the RNA polymerase and is unique to eukaryotes (reviewed in Boeger et al., 2005; Woychik and Hampsey, 2002). The RNA polymerase can be regarded as the core of the machinery because it is the platform upon which all the components are assembled. The transcription machinery is composed of a total of nearly 60 subunits and a mass of approximately 3 million Daltons (reviewed in Boeger et al., 2005).

The transcription cycle can be divided mechanistically in three basic stages: initiation, elongation and termination (Figure 2B). During each of these stages, RNA Pol II associates transiently with many different factors. Initiation involves the binding of transcription factors and RNA Pol II to DNA sites adjacent to the start site of transcription (the promoter regions) and the onset of RNA synthesis. After productive initiation, transcription proceeds into elongation mode and this is accompanied by a partial disassembly of the initiation complex and the association with elongation factors, which enable efficient production of long transcripts. In this stage the polymerase moves 5' to 3' along the DNA and makes an RNA copy of the gene. Finally, the termination stage involves release of the transcripts from RNA polymerase and release of the polymerase from the DNA template, allowing recycling of the polymerase and preventing it from perturbing promoters of genes located downstream from the transcription unit (reviewed by Howe, 2002; Orphanides and Reinberg, 2002).

The largest of the 12 subunits of RNA Pol II possesses a unique domain, not related to regions of any known protein, at its carboxyl terminus, termed the carboxyl-terminal domain (CTD) (Figure 2A). The CTD consists of heptapeptide repeats with the consensus Tyr^1-Ser^2-Pro^3-Thr^4-Ser^5-Pro^6-Ser^7 (Corden et al., 1985), which has been conserved through evolution (Allison et al., 1988; Barron-Casella and Corden, 1992). The number of repeats varies among species, ranging from 26-27 in yeast (Allison et al., 1985) to 52 in mammals (Corden et al.,

1985). The CTD is rich in potential phosphoacceptor amino acid residues and undergoes a cycle of phosphorylation and dephosphorylation during the transcription cycle (Figure 2B) (for reviews see Howe, 2002; Kobor and Greenblatt, 2002).

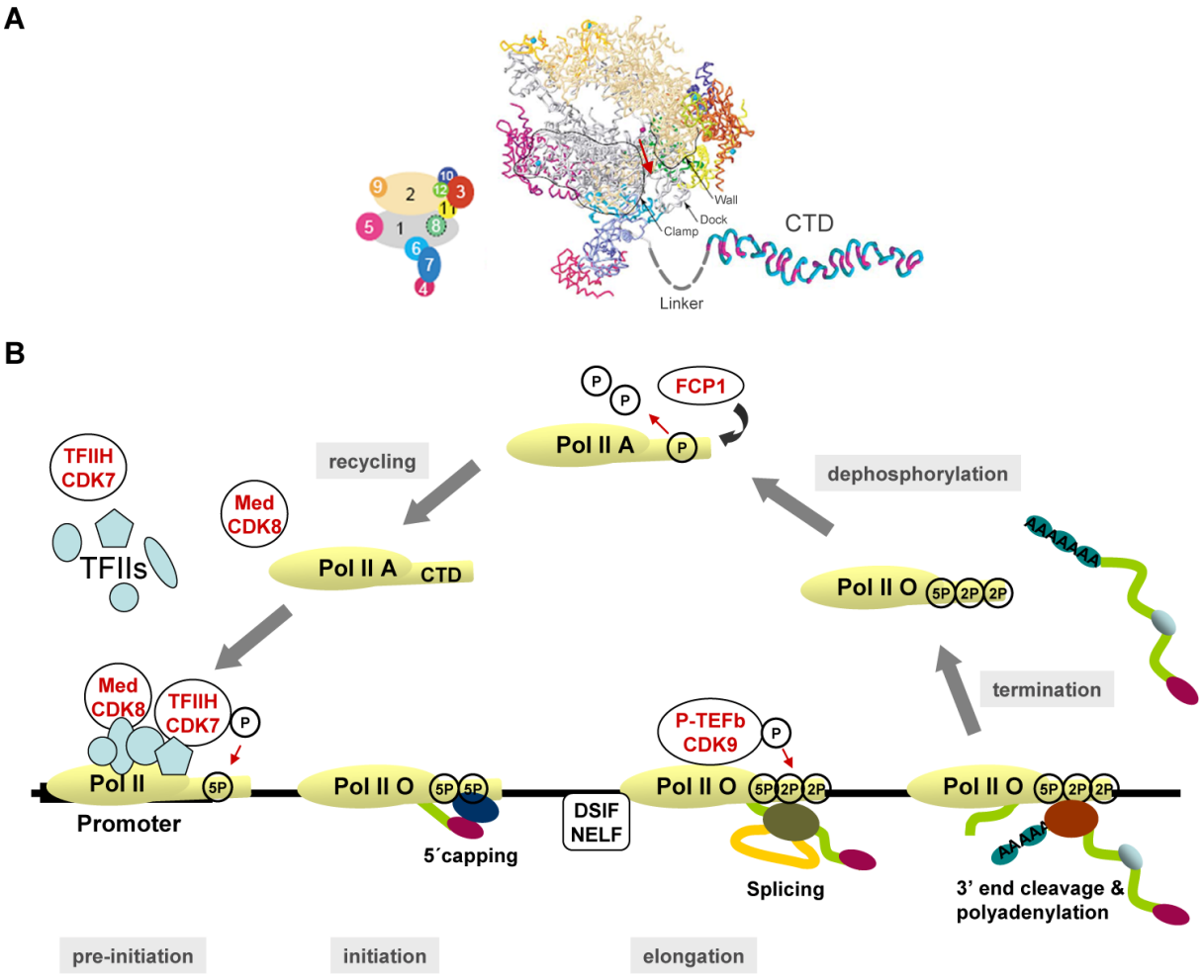


Figure 2 - RNA Polymerase II and the transcription cycle. (A) Model of the complete 12-subunit RNA Polymerase II (Armache et al., 2003; Bushnell and Kornberg, 2003). The view is from the top. Cyan spheres and a pink sphere depict eight zinc ions and an active site magnesium ion, respectively. The red arrow indicates the direction of RNA exit. A key to subunit colour is shown on the left with subunits numbered 1-12. The carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II (RNA Pol II LS, subunit 1) is shown in a β-spiral model (Meinhart and Cramer, 2004). The CTD is connected to the structured core of RNA Pol II with the mobile linker of approximalely 90 amino acid residues in length. Adapted from (Cramer, 2004; Meinhart and Cramer, 2004). **(B)** CTD phosphorylation and the transcription cycle. After recycling the "free" RNA Pol II core enzyme is not phosphorylated on the CTD. It may assemble with coactivators such as the mediator complex (Med) thus forming a holoenzyme. Premature CTD phosphorylation by CDK8 prevents the assembly of RNA Pol II on the promotor. For pre-initiation the unphosphorylated RNA Pol II core or holoenzyme assembles onto the promoter sequences with general transcription factors (TFIIs) thus forming a pre-inititiation complex of

transcription. The CTD is phosphorylated on serines at position 5 (5P) by the CDK7 subunit of TFIIH. Initiation follows and transcription begins. The phosphorylated CTD recruits the capping enzymes (*dark blue oval*) and the nascent transcript (*geen line*) is capped at its 5' end (*purple oval*). During elongation the CTD is phosphorylated on serines at position 2 (2P) by the CDK9 subunit of the positive transcription elongation factor (P-TEFb) which is required to remove the block opposed by the DSIF/NELF factors and elongate transcription. The phosphorylated CTD recruits the splicing machinery (*green oval*) to remove introns (*yellow line*) and finally recruits the cleavage and polyadenylation factors (*red oval*) that cleave the transcript and add a poly(A) tail at its 3' end. This step signals transcription to terminate and RNA Pol II falls off its DNA template. The resulting mRNA is exported to the cytoplasm. The blue ovals on the mRNA molecules represent proteins that are bound to the mRNA forming a messenger ribonucleoprotein particle (mRNP). To be recycled for another transcription round, RNA Pol II is dephosphorylated by the FCP1 CTD phosphatase. Adapted from (Palancade and Bensaude, 2003).

RNA Pol II with a hypophosphorylated CTD is recruited to promoters (Lu et al., 1991). Shortly after transcription begins, the CTD becomes phosphorylated on Ser^5. This modification signals the polymerase to clear the promoter and shift into an elongation mode. During the elongation stage of the transcription cycle, phosphorylation of Ser^2 predominates (Komarnitsky et al., 2000). During or after transcription termination, the CTD is dephosphorylated by CTD phosphatases, the best studied of which is FCP1, resulting in recycling of the largest subunit of RNA Pol II (Figure 2B) (Cho et al., 1999; Kobor et al., 1999).

The high-resolution crystal structure of RNA Pol II has provided detailed insight into how a catalytically active polymerase is structured and has revealed important aspects of its function (reviewed in Boeger et al., 2005; Woychik and Hampsey, 2002). It is now known that the RNA leaves the polymerase active center cleft via the exit tunnel and then disengages from the enzyme surface. When the RNA reaches lengths if 26 and 29 nt, its 5' end associates with RNA Pol II at the base of the dock domain (Andrecka et al., 2008). Interestingly, the CTD was shown to be placed near the exit tunnel (Cramer et al., 2001; Douziech et al., 1999) and the RNA was shown to extend toward the linker connecting to the CTD (Andrecka et al., 2008). These structural observations are consistent with the coupling of transcription with the processing events and to the role of the CTD as the major key player in this coupling, as will be discussed in detail in section 3.

2.2 Pre-mRNA processing

In eukaryotes, before a gene transcript is ready to be transported out of the nucleus, it has to undergo three major processing events to produce the fully translatable mRNA. These comprise the acquisition of a cap structure at the 5' terminus, the splicing out of non-coding intervening sequences (introns) from the pre-mRNA, and the generation of a 3' end, usually modified by the addition of a poly(A) tail. Some transcripts are also subject to an additional processing event termed RNA editing.

Capping

The 5' end of the pre-mRNA is modified soon after its synthesis, when the transcript is about 25–30 nucleotides in length, by the addition of 7-methyl guanosine, usually called the cap structure (Coppola et al., 1983; Moteki and Price, 2002; Rasmussen and Lis, 1993). The addition of the cap structure is carried out by three enzymatic activities in a process known as capping (for a review see Gu and Lima, 2005). First, RNA 5'-triphosphatase hydrolyses the triphosphate of the first nucleotide to a diphosphate. Then, RNA guanylyltransferase catalyses the fusion of a GMP moiety from GTP to the diphosphate end of the pre-mRNA. To finalize the cap structure, the RNA-(guanine-7) methyltransferase adds a methyl group to the N7 position of the transferred GMP to form the m7G(5')ppp(5')N cap. In metazoans, both the RNA 5'-triphosphatase and the RNA guanylyltransferase activities are part of the same polypeptide at the N-terminus and C-terminus, respectively, called the capping enzyme. In *Saccharomyces cerevisiae (S. cerevisiae)*, the capping enzyme consists of two polypeptides, RNA triphosphatase (Cet1) and RNA guanylyltransferase (Ceg1), which form a heterodimer (reviewed by Shuman, 2001).

The initial cap structure is recognised by the cap binding complex (CBC), which contains two proteins, CBP20 and CBP80. The cap structure bound to CBC is believed to play a major role in the stabilization of the mRNA, since it provides an obstacle for 5' to 3' exonucleases. Upon export through the nuclear pore complex, the nuclear cap binding proteins are replaced by the cytoplasmic translation initiation factor, eIF-4E. In the cytoplasm, cap bound to eIF-4E and other translation initiation factors enhances translation by promoting the engagement of the ribosomal subunits with the mRNA (see Mitchell and Tollervey, 2001).

Splicing

In the majority of mammalian genes the coding sequences are interrupted by non-coding intervening sequences (introns) that must be removed from the primary transcript (pre-mRNA) in order to generate a mature functional mRNA for translation (for a review see Sharp, 2005). In humans, a typical pre-mRNA contains seven or eight introns (Lander et al., 2001). The precise removal of the introns and joining of the coding sequences (exons), by a process known as splicing, is a critical step in gene expression. The splicing reaction is carried out by a large macromolecular complex termed the spliceosome. This is a highly dynamic and complex molecular machine whose composition and structure undergoes several rearrangements during each cycle of splicing. To date, two spliceosomes of unique composition have been characterised. The first one to be identified, now called U2-dependent or major spliceosome, is found in all eukaryotes and splices the most commonly encountered class of introns (the U2-type introns). It is composed of five small nuclear ribonucleoprotein particles (snRNPs; U1, U2, U4, U5 and U6), each of them containing a single uridine-rich small nuclear RNA (snRNA) that is associated to a common core of Sm proteins and other proteins specific of each particular snRNP. This spliceosome contains also a large number of non-snRNP proteins, known as splicing factors that exert auxiliary functions in the splicing reaction (reviewed by Jurica and Moore, 2003). The major class of non-snRNP proteins are the serine-arginine (SR) proteins, characterised by the presence of one or two RNA-binding domains and a carboxyl-terminal SR-rich domain (RS-domain) (Graveley, 2000). A less abundant spliceosome was discovered more recently in a subset of eukaryotes and is termed U12-dependent or minor spliceosome. This consists of the U11, U12, U5 and U4atac/U6atac snRNPs and an unknown number of non-snRNP proteins. Members of the SR protein family have been shown to function also in the U12-dependent splicing. Both the major and the minor spliceosomes coexist in eukaryotic cells and pre-mRNAs containing both types of introns serve as substrates for both splicing machineries. Although U12-type introns represent less than 1% of all introns present in human cells, they are found in genes carrying out essential cellular functions (for a review see Will and Luhrmann, 2005). It was recently reported that the two splicing pathways (minor and major) are spatially separated in the cell. *In situ* hybridisation on tissue sections from adult zebrafish and in mammalian cultured cells revealed a perinuclear and cytoplasmic staining for the U12 and U6atac, compared to the well-established nuclear distribution of the U2 snRNA of the major spliceosome. Functional analysis revealed that the cytoplasmic minor spliceosome is active and pointed to a role in cell

proliferation (Konig et al., 2007). U2-dependent splicing was also reported in the cytoplasm of platelets and dendritic cells but the full functional and physiological relevance of cytoplasmic splicing is not entirely clear yet (Denis et al., 2005; Glanzer et al., 2005).

The spliceosome acts through RNA-RNA, RNA-protein and protein-protein interactions to recognise the exon-intron junctions and catalyse removal of the introns and the joining of the exons in the correct order (see Faustino and Cooper, 2003). There are four short sequences that define an intron: the exon-intron junction at the 5' and 3' ends of the introns (5' and 3' splice sites), the branch point sequence, which includes an adenosine residue and the polypyrimidine tract (Py tract) that precedes the 3' splice site (see Ast, 2004). Different splice sites and branch point sequences are found in U2- and U12- type introns. U2- type introns of higher eukaryotes have the sequence AG/GURAGU at the 5' splice site, CURACU at the branch point sequence and YAG/G at the 3' splice site (where / denotes an exon-intron boundary, A is the branchpoint adenosine, R is a purine and Y a pyrimidine). The vast majority of U2-type introns have GU and AG dinucleotides at their 5' and 3' ends, respectively (Figure 3A) (Will and Luhrmann, 2005). By comparison, consensus sequences delineating the U12-type 5' splice site –A(or G)UAUCCUUU– and branch point sequence –UCCUUAACU– are longer and more tightly constrained. The 3' splice sites of U12-type introns are typically denoted by YAC/ or YAG/, but various other dinucleotides can serve as U12- type 3' splice sites. Similarly to U2-type introns present in yeast, U12-type introns lack a polypyrimidine tract. Because the first set of U12-type introns identified contained the dinucleotides AT and AC at their 5' and 3' ends, respectively, they were originally referred to as ATAC introns. Further studies, however, revealed that some U2-type introns also end in AT-AC and a large number of U12-type introns also end in GT-AG (see Will and Luhrmann, 2005). The splicing consensus sequences are short and weakly conserved, thus the splice site recognition requires additional RNA sequences that can be present in both introns or exons and function as splicing enhancers or repressors modulating the ability of the spliceosome to recognise nearby splice sites (reviewed by Matlin et al., 2005). In most organisms, exons or introns are recognised as a unit in "exon definition" for short exons or "intron definition" for short introns (Robberson et al., 1990).

In chemical terms, the splicing reaction involves two *trans*-esterification reactions and the U2-type and U12-type introns appear to be removed by an identical mechanism. First the 2'OH group of the branch point adenosine acts as a nucleophile to attack the phosphate at the 5' splice site, and *trans*-esterification results in a free 5' exon and a lariat-shaped molecule

consisting of the intron sequences and the 3' exon. In the second reaction, the 3'OH group of the free 5' exon attacks the phosphate at the 3' end of the intron. The subsequent *trans*-esterification results in the fusion of the two exon sequences and the release of the lariat-shaped intron (see Proudfoot et al., 2002; Will and Luhrmann, 2005).

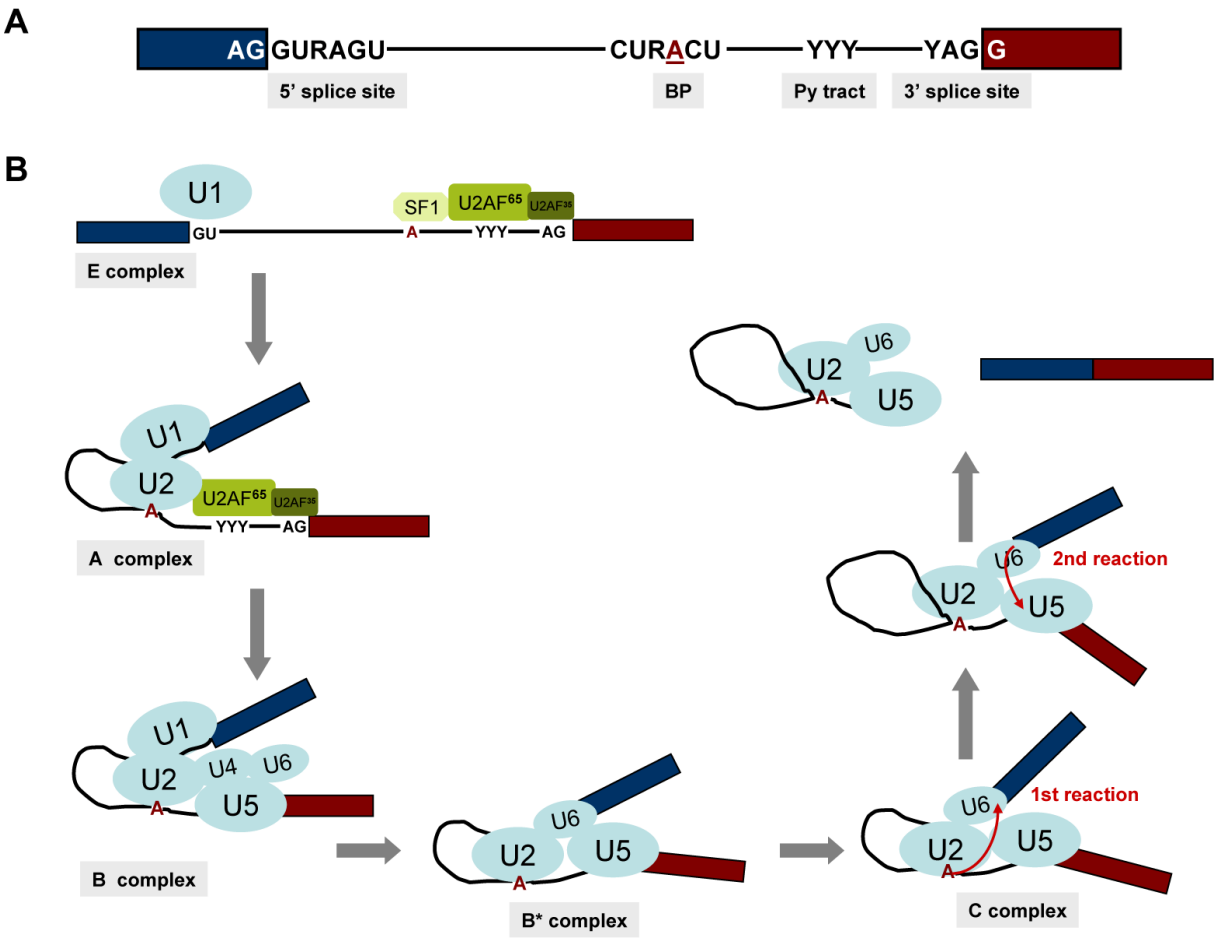


Figure 3 - Splicing of U2-type introns. (A) Schematic representation of a U2-type intron with its splicing consensus sequences; 5' and 3' splice sites, branch point (*BP*) and polypyrimidine (*Py*) tract. The intron is flanked by two exons (*blue and red*), the branching adenosine is underlined and highlighted in red, R is a purine and Y a pyrimidine. **(B)** Simplified view of the spliceosome cycle for U2-type introns. Only the snRNPs U1, U2, U4, U5 and U6 and the splicing factors SF1, $U2AF^{65}$ and $U2AF^{35}$ are shown. Different spliceosome conformations can be found at specific time points and purified as stable (E, A, B and C) complexes. Splicing proceeds through two *trans*-esterification reactions (*red arrows*) within the active spliceosome complex. After the second splicing reaction the mRNA is released, the post-spliceosomal complex containing the excised intron and the U2, U5 and U6 snRNPs disassembles and the snRNPs are recycled for new rounds of splicing (see text for more details) (Ast, 2004; Jurica and Moore, 2003; Will and Luhrmann, 2005).

The dissection of the splicing reaction through *in vitro* studies gave rise to the knowledge we have today about the spliceosome cycle for U2-type introns. In these experiments, splicing is uncoupled from other pre-mRNA processing events and the spliceosome assembly proceeds through a stepwise series of assembly events giving rise to short-lived intermediate stages, named E, A, B and C complexes (Figure 3B).

The initial events of spliceosome assembly require interaction of the U1 snRNA with the 5' splice site. In higher eukaryotes, the 3' splice site and adjacent Py tract are identified through interactions with U2 snRNP auxiliary factor (U2AF) composed of a large subunit of 65 kDa ($U2AF^{65}$) and a small subunit of 35 kDa ($U2AF^{35}$). $U2AF^{65}$ is an essential splicing factor that binds to the polypyrimidine tract and contacts also the branch point (Guth et al., 2001; Kent et al., 2003), while $U2AF^{35}$ binds to the conserved AG dinucleotide at the 3' splice site and is dispensable for splicing of some introns that contain "strong" polypyrimidine tracts (Pacheco et al., 2006; Pacheco et al., 2004). The branch point is also specifically recognised by the splicing factor 1 / mammalian branch point binding protein (SF1/mBBP), in a cooperative binding with $U2AF^{65}$ (Selenko et al., 2003). These ATP independent events lead to the formation of an E (early) complex that commits the pre-mRNA to the splicing pathway. These interactions help the recruitment of the U2 snRNP to the branch point in an ATP dependent manner (Sanford and Caceres, 2004). U2 snRNA base pairs with the branch point sequences generating the pre-spliceosome or A complex. Several proteins of the SR family are thought to mediate interactions between adjacent 5' and 3' splice sites, both across the intron and over the exon helping to determine the correct splice site selection and to stabilize the A complex (Hertel and Graveley, 2005; Shen and Green, 2004). This complex recruits the pre-formed U4/U6.U5 tri-snRNP to form the mature spliceosome or B complex. The U4 snRNA does not directly interact with the pre-mRNA but plays an essential role in bringing U5 and U6 into the spliceosome. At this step, the U1 base pairing interaction with the 5' splice site is replaced by a similar interaction involving the U6 snRNA, while U5 binds to sequences in the 3' exon, thus bringing the two exons closer together. Extensive conformational changes promote the dissociation of U1 and U4 from the complex and the spliceosome is activated (B* complex) and capable of catalyzing the first splicing reaction whereby the branch point adenosine is connected to the 5' end of the intron, which is cleaved from the upstream exon. The intron is now in a lariat configuration and the C complex is formed. More conformational changes are required for the second *trans*-esterification reaction to occur, in which the 3' end of the upstream exon is joined with the 5'

end of the downstream exon, cleaving the intron from the 3' end splice site. After this final step, the spliced mRNA is released from the post-spliceosomal complex formed by the lariat intron and the snRNPs, which must disassemble and recycle for another round of splicing (Figure 3B) (reviewed by Hastings and Krainer, 2001; Jurica and Moore, 2003). Assembly of the U12-dependent spliceosome is similar to that of the U2-dependent, with the U11, U12 and U4atac/U6atac being functionally similar to the U1, U2 and U4/U6. The major difference occurs at the earliest assembly step in which the U11 and U12 snRNPs form a highly stable di-snRNP that binds cooperatively to the 5' splice site and branch point, which is equivalent to the A complex of the U2-dependent spliceosome (reviewed by Will and Luhrmann, 2005).

While many introns are constitutively excised during splicing, other introns and exons can either be included or excluded from the mature mRNA, in different combinations in a process called alternative pre-mRNA splicing (for reviews see Blencowe, 2006; Matlin et al., 2005). By this process a single pre-mRNA yields different mRNAs leading to the production of different proteins creating proteome diversity. It has been estimated that more than 50% of all human genes are, at some stage, expressed by alternative splicing (see Modrek and Lee, 2002).

3' end processing

The 3' end processing of the transcripts consists in a two-step reaction involving site-specific endonucleolitic cleavage of the pre-mRNA and synthesis, at the newly generated 3' product, of a poly(A) tail of 70–90 adenosine residues in yeast (*S. cerevisiae*) and 200–250 adenosine residues in higher eukaryotes. The two steps of the reaction (cleavage and polyadenylation) are considered to be tightly coupled, as cleaved 3' ends that were not polyadenylated have not been observed *in vivo*. With the exception of replication-dependent histone transcripts (in higher eukaryotes), that possess a stem-loop structure at their 3' end, all protein encoding mRNAs contain a poly(A) tail (see Proudfoot et al., 2002; Zhao et al., 1999). The poly(A) tail is an important determinant of mRNA metabolism and function, largely through the action of cellular poly(A) tail-binding proteins (PABPs) (for a review see Mangus et al., 2003). PABPs have a minimal binding site of 11–12 adenosines and, as a consequence, a single poly(A) tail is bound by multiple PABP molecules (Mangus et al., 2003). While higher eukaryotes have multiple PABPs, including several cytoplasmic forms (PABPC) and one nuclear (PABPN1), yeast cells have a predominantly cytoplasmic poly(A)-binding protein (Pab1p) although cell fractionation indicates that some Pab1p is also nuclear (Zhao et

al., 1999). Orthologs of PABPN1 were not identified in yeast. However, recent studies indicate that the yeast heterogeneous nuclear ribonucleoprotein (hnRNP) Nab2p is a strong candidate for a yeast equivalent of PABPN1 function, at least for a subset of mRNAs (Hector et al., 2002).

In mammals, three elements define the core polyadenylation signal: the highly conserved hexanucleotide AAUAAA found 10 to 30 nucleotides upstream of the cleavage site; a less conserved U-rich or GU-rich element located downstream of the cleavage site and the cleavage site itself, which occurs predominantly after a CA dinucleotide and becomes the point of poly(A) addition being generally referred as the poly(A) site (Figure 4A) (see Gilmartin, 2005).

The cleavage and polyadenylation machinery has been extensively studied *in vitro*. Because *in vitro* it is possible to uncouple the cleavage and polyadenylation reactions, most of the factors involved in each reaction have been identified. In mammals, these include: the cleavage-polyadenylation-specificity factor (CPSF), required for both cleavage and polyadenylation; the cleavage stimulatory factor (CstF), necessary only for cleavage; two cleavage factors (CF I_m and CF II_m - the subscript m is used to differentiate the mammalian factor from the yeast one), required only for cleavage; the poly(A) polymerase (PAP), whose major role is in polyadenylation but is also required for cleavage and PABPN1 required for polyadenylation. In addition, the CTD of the largest subunit of RNA polymerase II (RNA Pol II LS) was also found to be required for efficient pre-mRNA cleavage (Figure 4B) (reviewed by Proudfoot, 2004; Zhao et al., 1999).

CPSF is a large protein complex containing subunits of 160, 100, 70 and 30 kDa (CPSF-160, CPSF- 100, CPSF- 70 and CPSF- 30, respectively) and has been shown to recognise the AAUAAA signal through CPSF-160. CstF consists of three polypeptides of 77, 64 and 50 kDa and binds to the GU – rich sequence downstream to the site of cleavage through its 64 kDa subunit. CF I_m consists of three subunits that are able to interact directly with a pre-mRNA substrate. CF II_m, only recently partially purified, can be subdivided into two fractions of which only fraction CF IIA_m appears to be essential for the cleavage reaction. PAP is present in at least two enzymatically active forms obtained by alternative splicing, PAP I (77 kDa) and PAP II (82 kDa), differing only at the carboxyl terminus. CPSF and PAP are sufficient for poly(A) addition to a pre-cleaved RNA substrate. However, to achieve rapid elongation and to control the length of the poly(A) tail PABPN1 is essential (Proudfoot and O'Sullivan, 2002; Zhao et al., 1999). PABPN1 is a 33 kDa protein that binds specifically to

the poly(A) tail via two RNA binding domains (Kuhn et al., 2003; Tavanez et al., 2005), forming spherical protein-RNA complexes with a diameter of 21 nm (Keller et al., 2000).

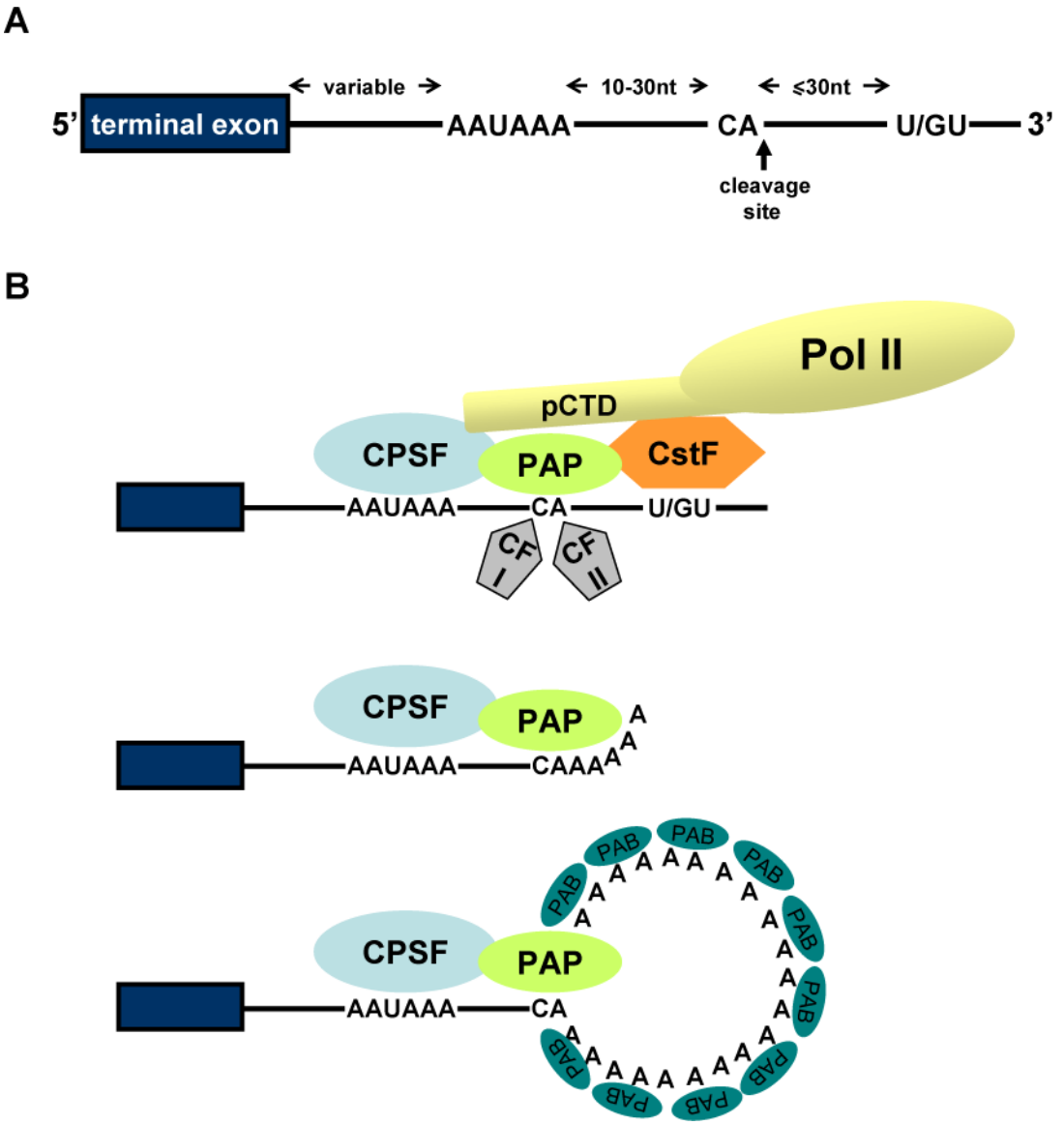


Figure 4 - Mammalian pre-mRNA 3' end processing. (A) Schematic representation of the mammalian poly(A) site with conserved sequence elements and relative distances between them. Adapted from (Gilmartin, 2005). **(B)** Simplified view of the cleavage and polyadenylation reactions. Direct cleavage of the mRNA requires CPSF, CstF, two additional cleavage factors (CF I and CF II), PAP and the phosphorylated CTD (pCTD) of RNA Polymerase II. PAP together with CPSF, directs poly(A) addition. PABPN1 (PAB) binds the growing poly(A) tail, greatly enhancing the efficiency of polyadenylation and forming 21 nm spherical particles (see text for details). Adapted from (Proudfoot et al., 2002).

In mammals the following model has been proposed for the 3' end processing reaction (Figure 4B). The initial step is probably the recognition of the signals on the pre-mRNA by CPSF and CstF in a process assisted by CF I_m. It is not known when the precise cleavage site is chosen, but the CPSF-CstF interaction defines the region in which it must lie. The formation of a cleavage-competent complex requires the additional recruitment of CF II_m and PAP and there is no indication of a reorganisation of the complex before catalysis, as occurs during the splicing reaction (Zhao et al., 1999). The nature of the endoribonuclease activity that is responsible for the cleavage reaction has remained elusive for a long time but recently two studies provided evidence that it might be the CPSF-73 subunit (Callebaut et al., 2002; Ryan et al., 2004). The CTD of RNA Pol II LS enhances the cleavage reaction, likely via interactions with CPSF and CstF (Hirose and Manley, 1998; McCracken et al., 1997b). The cleavage reaction is followed by polyadenylation. In this step PAP is recruited to the AAUAAA-containing substrate thought its interaction with CPSF and slow polymerization starts. After a binding site for PABPN1 has been created, the elongation proceeds at a faster speed until the poly(A) tail reaches approximately 200 adenosine residues. PABPN1 forms a spherical protein-RNA complexe with a 21 nm diameter (Keller et al., 2000), whose function remains unclear, although there has been some speculation that it might be related with measuring the length of the poly(A) tail (Kuhn and Wahle, 2004). When the Poly(A) tail reaches the right size processive elongation switches to a slow mode again and termination of polymerization is probably coupled with the release of CPSF and PAP from the finished product, although the mechanism is still not understood (see Zhao et al., 1999).

In *S. cerevisiae* the *cis* elements for 3' end formation are poorly defined, though three redundant and degenerate sequences are generally identifiable: the polyadenylation-cleavage site (PyA_n), an "efficiency element" (UAUAUA, or repeats thereof) located a variable number of nucleotides 5' to the poly(A) site, and a "positioning element" (AAUAAA, AAAAAA, and related sequences) located approximately 20 nucleotides 5' to the poly(A) site (reviewed by Zhao et al., 1999). The 3' end processing machinery in yeast includes the cleavage and polyadenylation factor IA (CFIA), cleavage and polyadenylation factor IB (CFIB), cleavage and polyadenylation factor (CPF), poly(A) polymerase (Pap1p), poly(A)-specific nuclease (PAN) and the poly(A)-binding proteins Pab1p and Nab2p (Proudfoot and O'Sullivan, 2002; Zhao et al., 1999). Since many of the yeast factors have only recently been characterised, less is known about how they interact with the RNA precursor and with each other. CFIA contains subunits homologous to those of mammalian CstF and CF II_m while CPF contains homologs

of CPSF as well as other mammalian cleavage and polyadenylation factors. During 3' end processing, multiple Pab1 proteins are loaded onto the poly(A) tail and are exported as part of the mensenger ribonucleoprotein particle (mRNP). One nuclear function of Pab1p is to modulate the overall length of the poly(A) tail, which is initially synthesized at a length greater than is found on newly exported mRNA. The multisubunit complex PAN is recruited by Pab1p and is thought to trim the poly(A) tail to the appropriate length (Dunn et al., 2005). The nuclear poly(A)-binding protein Nab2p is also required for poly(A) tail length control. Loss of NAB2 expression leads to hyperadenylation of the transcripts and although cells are viable in the absence of NAB2 expression when PAB1 is overexpressed, Pab1p fails to resolve the hyperadenylation defect indicating that Nab2p is essential for poly(A) tail length control *in vivo* (Hector et al., 2002).

Editing

In addition to capping, splicing and 3' end formation, another RNA processing reaction that affects the pre-mRNA is editing. This co- or post-transcriptional RNA processing reaction changes the nucleotide sequence of the RNA substrate by insertion or deletion of nucleotides or by the substitution of bases by modification (for reviews see Gott and Emeson, 2000; Schaub and Keller, 2002). The major type of RNA editing in the nucleus of higher eukaryotes is base modification and the best characterised reactions are hydrolytic deaminations with conversion of cytidine (C) to uridine (U) and adenosine (A) to inosine (I). The conversion of C to U in pre-mRNA is catalysed by cytidine deaminase acting on RNA (CDAR) complexes and conversion of A to I is catalysed by adenosine deaminase that act on RNA (ADAR) polipeptides (reviewed by Gerber and Keller, 2001). In mammals, two active ADAR enzymes with different substrate specificities have been characterised, ADAR1 and ADAR2 (Hartner et al., 2004; Higuchi et al., 2000). The first RNA-editing reaction discovered in mammals was the tissue-specific modification of apolipoprotein B mRNA (Powell et al., 1987). In this substrate, deamination of C to U changes a glutamine codon (CAA) to a stop codon (UAA), resulting in the translation of a shorter protein termed apoB48. In humans, such editing is restricted to the small intestine, and the unedited full-length protein (apoB100) is synthesized in the liver (see Gerber and Keller, 2001). This is one of the few examples in mammals of C to U changes in mRNA resulting from RNA editing. The A to I conversion which is read by the translation machinery as a G, is the most frequent editing reaction in higher eukaryots and has been described in different tissues but most of the known

targets are pre-mRNAs expressed in the central nervous system (see Schaub and Keller, 2002). Well studied editing substrates in the brain include subunits of the glutamate ion channel family or the serotonin 5-HT_{2C} receptor (Gerber and Keller, 2001). ADAR editing occurs with low selectivity on completely double-stranded RNA (dsRNA) or selectively on structured RNA, interrupted by bulges of loops (Bass, 2002). Because dsRNA structure can be formed by pairing of intronic and exonic sequences it was hypothesised that the speed of splicing could regulate the extent of editing. However, it was shown that at least for the glutamate receptor subunit B pre-mRNA the absence of editing prevents splicing of an intron adjacent to an editing site, which showed that editing can regulate splicing (Higuchi et al., 2000). It was also shown recently that the CTD of RNA pol II is required for efficient co-transcriptional editing of ADAR2 pre-mRNA (Laurencikiene et al., 2006) and that the CTD and ADAR2 function together to enforce the order of events that allows editing to precede splicing in the glutamate receptor subunit B pre-mRNA (Ryman et al., 2007).

So far only a few pre-mRNA substrates of editing have been identified, nevertheless, the impact of pre-mRNA editing in the regulation of gene expression can be significant. Editing can affect other pre-mRNA processing reactions such as splicing, alter the specificity of a codon, create new start and stop codons or even new open reading frames generating protein isoforms with different functions that could not be predicted from the corresponding genomic sequence (for reviews see Gott and Emeson, 2000; Schaub and Keller, 2002).

2.3 Nuclear export and translation

Cellular mRNA molecules produced in the nucleus must be exported to the cytoplasm to allow their translation into proteins. Many of the proteins synthesized in the cytoplasm have a role in nuclear metabolism. This division requires nucleo-cytoplasmic transport pathways that can rapidly and specifically transport the molecules from their site of synthesis to their site of use. Nuclear export of mRNA, like the exchange of all macromolecules between the nucleus and the cytoplasm, occurs through large structures embedded in the nuclear envelope called nuclear pore complexes (NPCs) (see Allen et al., 2000; Ryan and Wente, 2000). In higher eukaryotes, NPCs are composed of approximately 50-80 distinct proteins, collectively referred to as nucleoporins (Nups), which assemble into a supramolecular structure forming an aqueous channel across the nuclear envelope. A subset of these nucleoporins contains characteristic domains with multiple repeats of short peptide

sequences ending in the amino acids phenylalanine and glycine (FG). These FG repeats are believed to function as transient docking sites for nucleo-cytoplasmic transport factors. Other features of the NPC include the nuclear basket, nuclear and cytoplasmic rings, short cytoplasmic fibrils, and a central spoke complex (reviewed by Vasu and Forbes, 2001). Small molecules (≤ 40 kDa) can cross the NPC by passive diffusion, but most of the molecules that move between the nucleus and the cytoplasm are actively transported. Soluble proteins known as transport receptors are required for passage of the molecules to be transported (cargos) through the NPC by mediating interactions with the nucleoporins. The bulk of cellular nucleo-cytoplasmic transport is mediated by factors that belong to a single family of nuclear transport receptors termed "karyopherins" or importins/exportins. Different members of this protein family bind to distinct cargo molecules, or to adapter proteins that in turn bind cargo, and also share the ability to interact with specific nucleoporins. For example, U snRNA and tRNA require exportin 1/CRM1 and exportin-t, respectively. A karyopherin family member that functions in general mRNA export has not been identified (reviewed by Cullen, 2003).

Nuclear mRNA molecules exist as large RNA-protein complexes, thus the transport substrate recognised by the mRNA export machinery is the messenger ribonucleoprotein particle (mRNP). Despite intensive investigation, the molecular mechanisms of mRNA export are not completely clear yet. A general feature of mRNA export is that it is an active, receptor-mediated process. The mRNA export receptors are thought to recognise and bind to the mRNP-export cargoes either directly or indirectly (via adaptor proteins) and facilitate their translocation across the central channel of the NPC. On the cytoplasmic side of the NPC, the exported mRNP is released and the receptor returns to the nucleoplasm, without the cargo, to initiate additional rounds of export. Recent studies in yeast and in higher eukaryotes have led to the elucidation of an evolutionarily conserved pathway for the export of bulk mRNP to the cytoplasm. The best candidate for an mRNP export receptor in a heterodimer composed of a large subunit, TAP or NXF1, and a small cofactor called p15 or NXT1. In yeast cells the TAP ortholog is Mex67p and the small cofactor is Mtr2p. NXF1 belongs to a conserved family of nuclear export factor proteins (NXF) that are distinct from the karyopherin family of transport receptors but associate with NPC and have the ability to shuttle. NXT1 is conserved in metazoan but does not show any obvious sequence similarity to Mtr2, being nevertheless functionally analogous. Mtr2p is required for the binding of Mex67p to NPCs and NXT1 is also essential for NXF1 association with nucleoporins and recruitment to NPCs *in vivo* (reviewed by Erkmann and Kutay, 2004; Stewart, 2007; Stutz and Izaurralde, 2003).

Mex67p was first detected in yeast via a synthetic lethal screen with Nup85p, a nucleoporin involved in mRNA export. Mutants of the *MEX67* gene accumulate poly(A)$^+$ RNA in the nucleus (Santos-Rosa et al., 1998; Segref et al., 1997). Evidence that NXF1 (originally called TAP) is involved in mRNA export in metazoans was first provided by studies of a host factor that facilitates export of type D retroviral RNAs by binding directly to the constitutive transport element (CTE) of these RNAs (Bray et al., 1994; Gruter et al., 1998). Subsequent work showed that NXF1-NXT1 (or TAP-p15) stimulates export of cellular mRNAs, confirming its role in mRNA export (Braun et al., 2001; Guzik et al., 2001; Katahira et al., 1999).

Although NXF1 and Mex67p exhibit a general RNA binding affinity *in vitro*, their association with nuclear mRNAs is thought to be mediated by protein-protein interactions (see Stutz and Izaurralde, 2003). Proteins that might act as adapters between mRNAs and the heterodimer have been identified through genetic and biochemical approaches. In *S. cerevisiae*, Yra1p was identified as an mRNA export adapter for Mex67p. *YRA1* gene encodes an RNA binding protein, which causes nuclear accumulation of poly(A)$^+$ mRNA when it is deleted (Strasser and Hurt, 2000; Stutz et al., 2000; Zenklusen et al., 2001). Proteins homologous to Yra1p have been identified in human, mouse, *C. elegans*, *Xenopus* and *Drosophila*, and belong to an evolutionary conserved protein family called RNA export factor binding proteins or REFs. These proteins bind directly to both RNA and NXF1 (or Mex67p) (see Erkmann and Kutay, 2004). In *Xenopus* oocytes, antibodies specific to REFs that prevent their interaction with RNA reduce the export rate of mRNA (Rodrigues et al., 2001) and microinjection of recombinant REF1 stimulates the export of mRNA that was inefficiently exported (Rodrigues et al., 2001; Zhou et al., 2000). Nevertheless, studies in *Drosophila* cells indicate that REF might contribute to but is not essential for mRNA export (Gatfield and Izaurralde, 2002). REFs are also dispensable for mRNA export in *C. elegans* (Longman et al., 2003). This raises the possibility that other adaptor protein(s) can bridge the interaction between NXF1-NXT1 (or TAP-p15) and cellular mRNAs in metazoan. Exactly how NXF1-NXT1 and the adaptor proteins are recruited to mRNAs remains uncertain but there is increasing evidence that it may be coupled to the synthesis and/or post-transcriptional processing of mRNAs (discussed in detail in section 3).

Once in the cytoplasm, mRNA molecules carry the genetic information to the protein-synthesizing machinery, where it is used to define the amino acid sequence of the polypeptides, and ultimately the structure and function of proteins. However, the translation

of mRNA is only the first step in the formation of a protein. The polypeptide chain is in most cases subjected to post-translational modifications and must fold into the appropriate three-dimensional conformation to achieve its active form (see Hausman, 2004). Translation takes place on the ribosome, a large ribonucleoprotein particle that consists of two subunits, designated 40S and 60S in eukaryotes, and involves also a large number of additional protein factors. Translation of mammalian mRNAs occurs by a scanning mechanism, in which the 40S ribosomal subunit, primed for initiation by the binding of several initiation factors, is loaded on the mRNA immediately downstream of the 5' cap and scans the mRNA in the 5' to 3' direction. Upon recognition of the initiation codon (the first AUG triplet) via base pairing with the Met-tRNA(i) anticodon, scanning ceases and the 60S subunit joins completing the initiation process (reviewed by Jackson, 2005).

It is now accepted that some proteins recruited to transcripts during their biogenesis in the nucleus, accompany the mRNAs to the cytoplasm and may influence their fate and ultimately their translation (see Dreyfuss et al., 2002; Kuersten and Goodwin, 2005). Furthermore, newly exported mRNPs have a protein composition distinct from mRNPs that have undergone the initial round of translation. For example, during this initial round the nuclear cap-binding complex (CBC), a dimmer composed of CBP20 and CBP80, is replaced by the cytoplasmic translation initiation factor, eIF-4F. The interaction between the cap-binding subunit of eIF-4F (eIF-4E) and the poly(A)-binding protein (PABP) will circularize the mRNP resulting in a translationally competent complex. The "closed loop" conformation of the mRNP may promote translational competence by enabling the recycling of ribosomes from the 3' to the 5' end of the mRNA (see Mitchell and Tollervey, 2001). The first round of translation has also been proposed to constitute a final quality control step in mRNA biogenesis allowing the identification of messages that contain premature termination codons (Kuersten and Goodwin, 2005). Recently it was proposed that this pioneer round of translation occurs when the mRNA is still in the nucleus challenging the prevailing view that translation is limited to cytoplasmic ribosomes (Iborra et al., 2004; Iborra et al., 2001).

3. Coupling between steps of gene expression

Studies over the past few years have revealed that virtually all of the steps in gene expression are extensively coupled to one another (see for example Bentley, 1999; Hirose and Manley, 2000; Kornblihtt et al., 2004; Proudfoot et al., 2002; Zorio and Bentley, 2004). Recent findings revealed that a highly interconnected system encompassing gene transcription, pre-mRNA processing, mRNA export and cytoplasmic mRNA translation and turnover operate in eukaryotic cells (reviewed by Maniatis and Reed, 2002). The major pre-mRNA processing steps – 5' capping, splicing and 3' end processing – are coupled to transcription (reviewed by Proudfoot et al., 2002; Zorio and Bentley, 2004). This functional coupling is achieved by physical contacts between the protein machines that perform transcription and pre-mRNA processing. The link between these protein machines is the carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II (RNA Pol II LS). The assembly of different protein complexes on the CTD can function by increasing the local concentration of processing factors, by coupling the rate of transcription with the assembly of specific RNA-protein complexes, and by allosterically activating and inhibiting the same complexes (Bentley, 2005).

The first evidence that the CTD is important for pre-mRNA processing was provided by Bentley and colleagues (McCracken et al., 1997a; McCracken et al., 1997b), who demonstrated that deletion of the CTD prevents efficient capping, splicing and 3′ end cleavage *in vivo*. Coupling among different pre-mRNA processing steps seems also to be critical for gene expression (reviewed by Maniatis and Reed, 2002; Proudfoot et al., 2002). The existence of interactions between capping and splicing, as well as interactions between splicing and 3' end processing in conjunction with the recruitment of the splicing machinery to internal exons in the pre-mRNA, play an important role in ensuring that the correct splice sites are recognised and engaged in the splicing reaction and that the transcripts are correctly capped and polyadenylated. It has long been suspected that RNA processing and export must be in some way connected to ensure that only the final mRNA product leaves the nucleus. Recent progress in identifying soluble factors important to mRNA trafficking is beginning to reveal the molecular basis for such a functional coupling (for reviews see Cullen, 2003; Reed, 2003; Reed and Cheng, 2005).

All the interactions described above represent a quality control mechanism to ensure that no individual processing reaction is omitted and that they are all timely, accurate and

efficient with the ultimate goal of producing a mature mRNA molecule (Maniatis and Reed, 2002).

3.1 Connecting transcription with pre-mRNA processing

The connector: the carboxyl-terminal domain of RNA polymerase II

The CTD of RNA Pol II LS contains multiple potential phosphorylation sites within its heptapeptide consensus sequence and undergoes dynamic changes in phosphorylation during the transcription cycle, which plays a significant role in coordinating its multiple activities (Meinhart et al., 2005; Phatnani and Greenleaf, 2006). The CTD is mostly hypophosphorylated (form IIa) when recruited to the promoter and becomes hyperphosphorylated (form IIo) during transcription elongation. When transcription initiates, the CTD is first phosphorylated on Ser^5 by CDK7, the TFIIH-associated kinase and later on the Ser^2 residues by CDK9, the kinase associated to the positive transcription elongation factor b (P-TEFb) (reviewed by Palancade and Bensaude, 2003). It was recently shown that transcribing RNA Pol II is also phosphorylated at Ser^7 (Chapman et al., 2007). This phosphorylation is required for expression of snRNA but not protein-coding genes (Egloff et al., 2007). Chromatin immunoprecipitation (ChIP) assays, using antibodies directed to the hypo- and the hyperphosphorylated forms of the CTD, have helped to map the evolution of CTD phosphorylation along a transcribed gene *in vivo*. In *S. cerevisiae,* Ser^5 phosphorylation is detected primarily at promoter regions dependent on TFIIH. In contrast, Ser^2 phosphorylation is seen only in coding regions (Komarnitsky et al., 2000). In the mammalian genes studied, RNA Pol II is more concentrated in the promoter-proximal regions than in the interior regions and different phosphorylation forms of the CTD are also associated with distinct regions. Ser^5 phosphorylation is concentrated near the promoter, while Ser^2 phosphorylation is observed throughout the gene (Cheng and Sharp, 2003).

How exactly the changes in CTD phosphorylation are regulated is not well understood. It has been shown that CTD conformational changes induced by the peptidyl-prolyl isomerase Pin1 affect both CTD kinase and phosphatase activities by remodelling their substrate (Xu et al., 2003). Although Pin1 plays a significant role in regulating RNA Pol II CTD structure and function is still not known if this enzyme affects the recruitment of mRNA processing factors. The CTD is also O-glycosylated in serine residues in a manner that is

mutually exclusive with serine phosphorylation, suggesting that there could be cross-talk between these two modifications (Comer and Hart, 2001).

The CTD interacts with many factors throughout the transcription cycle, including the Mediator complex that regulates transcription initiation, the capping enzymes and polyadenylation factors (see Figure 2B). Rather than carrying all these factors throughout the transcription cycle, the CTD interacts dynamically with each factor at the appropriate time. But how can such a simple sequence interact with so many targets in such a coordinate way? One hypothesis is that there is a CTD "code" that specifies the position of the RNA Pol II in the transcription cycle (see Buratowski, 2003). A lot of information can be encoded in the CTD simply by regulating the timing of phosphorylation events during the transcription cycle. If we consider all the possible configurations resulting from proline isomer patterns, the level of complexity of the "code" increases considerably. The CTD "code" could consist in a series of different phosphorylation and conformational changes that generate configurations specific for binding of particular factors. The complexity of the CTD "code" could be even higher if we consider all the other covalent modifications that have been reported for the CTD (including ubiquitination, glycosylation and phosphorylation of other residues within the repeat) and the nonconsensus repeats found in most organisms (Buratowski, 2003).

Transcription and capping

Capping is the best described example of a pre-mRNA processing reaction coupled to transcription. Evidence has accumulated over the last decade, showing that capping is facilitated by an association of the capping enzymes with the CTD (reviewed by Bentley, 2005; Howe, 2002; Zorio and Bentley, 2004). In the yeast *S. cerevisiae,* the RNA guanylyltransferase (Ceg1) associates *in vitro* with the phosphorylated form of the CTD (Cho et al., 1997) but it cannot form the covalent GMP intermediate unless the RNA triphosphatase subunit (Cet1) is also present (Cho et al., 1998). Therefore, the yeast Ceg1 is regulated by allosteric interactions with both the Cet1 and the CTD (Cho et al., 1998), which may reflect the need to coordinate both activities. ChIP experiments in *S. cerevisiae* demonstrated that Cet1 and Ceg1 crosslink to promoter regions of transcribing genes and the CTD kinase kin28 (CDK 7 in mammals) is required for binding (Komarnitsky et al., 2000; Rodriguez et al., 2000; Schroeder et al., 2000). Both the extent of CTD phosphorylation at Ser^5 of the heptad repeat and the binding of capping enzymes decreased as the polymerase moved from the 5' to the 3' end of genes. The CTD phosphatase, Fcp1, is required for dissociation of capping

enzymes from the elongation complex (Schroeder et al., 2000). The yeast methyltransferase (Abd1) crosslinked with promoter regions and coding regions further downstream (Komarnitsky et al., 2000; Schroeder et al., 2000). In mammals, the capping enzyme is able to directly interact with the phosphorylated CTD through the guanylyltransferase domain (Fong and Bentley, 2001; Ho et al., 1998; McCracken et al., 1997a) and the human methyltransferase binds to the complex of human capping enzyme and phosphorylated CTD (Pillutla et al., 1998). The human capping enzyme can associate with CTD peptides that are phosphorylated on either Ser^2 or Ser^5; however only the Ser^5 phosphorylated CTD peptides stimulate the capping reaction (Ho and Shuman, 1999). Binding of the capping enzyme to the CTD not only localises the enzyme but also allosterically regulates it, reducing the Km of the mammalian guanylyltransferase for GTP (Ho and Shuman, 1999).

There is now convincing evidence that the link between transcription and capping works both ways. Not only the capping reactions are influenced by the transcription machinery, as mentioned above, but the capping enzymes may also have an important role in regulating early steps in transcription (reviewed by Bentley, 2005). It is believed that during the formation of the transcription initiation complex, or soon after initiation, DRB sensitivity-inducing factor (DSIF) is recruited to the transcription complex. Additionally, after initiation of transcription, the negative elongation factor (NELF) is recruited through interaction with DSIF. This results in the arrest of the transcription complex before it enters into productive elongation. DSIF/NELF mediated arrest is then overcome by the positive transcription elongating factor, P-TEFb, and the associated protein kinase, CDK9, which phosphorylates both the CTD at Ser^2 and Spt5, the larger subunit of DSIF (see Figure 2B) (reviewed by Orphanides and Reinberg, 2002). In mammals and in the fission yeast *Schizosaccharomyces pombe* (*S. Pombe*), Spt5 interacts directly with the RNA triphosphatase and guanylyltransferase and stimulates RNA guanylylation (Pei and Shuman, 2002; Wen and Shatkin, 1999). The *S. cerevisiae* Abd1, stabilizes RNA Pol II on some promoters and both Abd1 and Ceg1 stimulate early elongation (Kim et al., 2004a; Schroeder et al., 2004), whereas Cet1 inhibits re-initiation (Myers et al., 2002). Stimulation of transcription by Abd1 occurs in a methylation-defective mutant and is therefore independent of capping itself (Schroeder et al., 2004). The human capping enzyme stimulates promoter escape by countering the negative elongation factor NELF, and capping enzyme recruitment is enhanced by direct binding to the elongation factor Spt5 (Mandal et al., 2004). In summary, capping enzymes have a previously unsuspected ability to manipulate early steps in transcription. In this way they may operate a

checkpoint to ensure timely capping of the nascent pre-mRNA before commitment to processive elongation of the transcript, ensuring that only properly capped RNA molecules are extended.

Transcription and splicing

The first evidence for a functional connection between transcription and splicing came from studies back in the 80's, showing that when transcription of an RNA Pol II gene was directed by an RNA Pol I or III promoter, splicing was inhibited (Sisodia et al., 1987; Smale and Tjian, 1985). An explanation for this early observation came many years later with several studies pointing to the CTD of RNA Pol II as the landing platform that allows the assembly of the processing machinery (reviewed by Hirose and Manley, 2000). Since the CTD domain is exclusive of RNA Pol II, the processing machinery could not be efficiently recruited to the chimaeric genes and splicing was impaired.

The idea that the CTD may function as a landing platform for the assembly of the splicing machinery was first proposed in 1993 by Greenleaf (Greenleaf, 1993). He presented a speculative model for co-transcriptional splice site selection in the pre-mRNA involving interactions between the negatively charged hyperphosphorylated CTD of RNA Pol II and the positively charged splicing factors with an RS-domain. In the following years several studies reported physical and functional interactions between the CTD and the splicing machinery (reviewed by Hirose and Manley, 2000). Splicing factors were shown to immunoprecipitate with the hyperphosphorylated RNA Pol II (RNA Pol IIO), but not with the hypophosphorylated polymerase (RNA Pol IIA), even in the absence of pre-mRNA (Kim et al., 1997; Mortillaro et al., 1996). Antibodies directed against the CTD and CTD peptides have been shown to inhibit splicing *in vitro* (Chabot et al., 1995; Yuryev and Corden, 1996) and the overexpression of phosphorylated CTD peptides inhibit splicing in cultured mammalian cells (Du and Warren, 1997). Evidence for a direct role for RNA Pol II in the splicing reaction independently of transcription came from *in vitro* reconstituted splicing assays in which RNA Pol II influences very early steps in the assembly of the spliceosome. This assays showed also that the phosphorylation status of the CTD has a dramatic effect in the splicing reaction, with RNA Pol IIO form strongly activating splicing and the RNA Pol IIA inhibiting the reaction (Hirose et al., 1999). Furthermore, experiments in which plasmid templates or pre-synthesized full-length pre-mRNAs were microinjected into *Xenopus* oocytes indicated that both co-and post-transcriptional splicing can occur *in vivo*, but only co-

transcriptional splicing is dependent on CTD phosphorylation (Bird et al., 2004). New insights into the mechanism by which the CTD functions in splicing come from recent *in vitro* experiments using a chimeric protein in which the CTD of RNA Pol II was fused to the C-terminus of the splicing factor SF2/ASF (ASF–CTD). Compared to SF2/ASF alone, ASF–CTD increased the reaction rate during the early stages of splicing but only when the CTD moiety was phosphorylated and the SF2/ASF moiety was bound to RNA (Millhouse and Manley, 2005).

A number of *in vivo* experiments gave valuable clues for a functional role of the CTD in coupling transcription to splicing. In a pioneer study, Bentley and co-worker's showed that RNA molecules transcribed in transient transfection assays by RNA Pol II with a truncated CTD containing only 5 repeats are less efficiently spliced than RNA molecules made by an RNA Pol II with a full-length CTD (McCracken et al., 1997b). This same CTD truncation was also shown to prevent the targeting of the splicing machinery to a transcription site in mammalian cells, suggesting a role for the CTD in the intranuclear targeting of splicing factors to transcription sites *in vivo* (Misteli and Spector, 1999). An analysis using RNA Pol II with different regions of the CTD truncated showed that different CTD regions can serve distinct functions in pre-mRNA processing, with the CTD carboxyl-terminus (including heptads 27-52) supporting capping, splicing and 3' end processing whereas the amino terminus (heptad repeats 1-15 or 1-25) supports only capping (Fong and Bentley, 2001). Further dissection of the CTD sequence showed that the C-terminal 10 amino acid motif that follows the 52 heptad repeats is also essential for high level transcription, splicing and poly(A) site cleavage (Fong et al., 2003). Efficient mRNA synthesis from a transiently transfected reporter gene requires between 16 and 25 heptad repeats from either the N- or C-terminal half of the CTD plus the C-terminal 10 amino acid motif (Fong et al., 2003).

The finding that promoter structure can influence alternative splice site selection extended the concept of a physical and functional coupling between transcription and splicing, suggesting additional levels of interplay (Cramer et al., 1999; Cramer et al., 1997). Using a series of human α-globin/fibronectin (FN) minigenes that include the alternatively spliced extra domain I (EDI) exon of FN and differing in the promoter driving their expression, it was demonstrated that the extent of EDI splicing is dependent on the promoter structure from which the transcript originated, independently of the promoter strength (Cramer et al., 1997). Furthermore, it was also shown that the promoter structure affected the responsiveness of this alternative splicing to the activation by the SR proteins SF2/ASF and 9G8 (Cramer et al.,

1999). Since the effect of these splicing factors requires the presence of an intact EDI exonic splicing enhancer (ESE) in order to stimulate EDI inclusion, it was suggested that the transcription machinery modulates the recruitment of specific SR proteins to exonic splicing enhancers (Cramer et al., 1999). Promoter control of alternative splicing has been extended to the cystic fibrosis transmembrane regulator (CD44), the calcitonin-gene-related peptide genes and the fibroblast growth factor receptor 2 genes (reviewed by Kornblihtt, 2005). Two possible mechanisms can explain the effect of promoters on alternative splicing choices (see Kornblihtt, 2005). The factor recruitment mechanism by which the promoter itself is responsible for recruiting splicing factors (such as SR proteins) to the site of transcription, possibly through transcription factors that bind the promoter or the transcriptional enhancers. This mechanism is supported by the observation that promoter structure affected the responsiveness of EDI alternative splicing to the activation by the SR proteins SF2/ASF and 9G8 (Cramer et al., 1999). In the other mechanism proposed, promoters can control alternative splicing via the regulation on RNA Pol II elongation rates or processivity. Studies of the effect of various transcriptional activators on the splicing of the FN EDI exon show that rapid, highly processive transcription favours EDI skipping, whereas slower, less processive transcription favours inclusion (Nogues et al., 2003). To explain this, a kinetic coupling model has been proposed (Caceres and Kornblihtt, 2002), where the transcript elongation rate determines the outcome of competing splicing reactions that occur co-transcriptionally. This has been directly shown *in vivo*, in human cells, using a mutant RNA Pol II with a lower elongation rate that it is able to affect alternative splicing of FN EDI exon and adenovirus E1a pre-mRNA (de la Mata et al., 2003). It was also shown in *S. cerevisiae* using a slow RNA Pol II mutant and inhibitors of RNA pol II elongation that changes in transcriptional elongation affected an alternatively spliced exon (Howe et al., 2003).

There is now evidence indicating that not only the transcriptional machinery can influence splicing, but the splicing components can also influence the efficiency of transcription, indicating that there is much more extensive coupling between these processes than previously anticipated (see Rosonina and Blencowe, 2002). It has been shown that snRNPs associate with the human transcription elongation factor TAT-SF1, forming a complex that can stimulate both transcription and splicing (Fong and Zhou, 2001). This effect is likely to be mediated through the binding of TAT-SF1 to elongation factor P-TEFb. Inclusion of splicing signals in the nascent transcript further stimulates transcription, supporting the notion that the recruitment of U snRNPs near the elongating polymerase is

important for transcription. Because the TAT-SF1-U snRNP complex also stimulates splicing *in vitro*, it may serve as a dual-function factor to couple transcription and splicing and to facilitate their reciprocal activation (Fong and Zhou, 2001). But what could be the functional relevance of the reciprocal coupling between transcription and splicing? A typical mammalian gene contains nine introns and each intron can range in length from hundreds to tens of thousands of bases (Zorio and Bentley, 2004). By stimulating the transcription machinery, assembly of splicing factors at nascent 5′ splice sites might therefore help RNA Pol II to transcribe through a large intron sequence. This could be achieved by facilitating the recruitment of P-TEFb that helps to maintain the phosphorylation status of the CTD as hyperphosphorylated which is associated with it's high processivity (Rosonina and Blencowe, 2002). Ultimately, this mutual stimulation of transcription and splicing ensures that they act as a close couple, thereby facilitating efficient gene expression.

Transcription and 3' end processing

The finding, back in the 80's, that a poly(A) site is required for termination of RNA Pol II transcription was the first clue for the existence of communication between the RNA processing and transcription machineries (Falck-Pedersen et al., 1985; Logan et al., 1987; Whitelaw and Proudfoot, 1986). Many years passed until the nature of this communication started to be elucidated. Meanwhile, other connections between 3' end processing and transcription started to be revealed, especially since the discovery of the CTD of RNA Pol II as a key player in connecting transcription and RNA processing. Experiments using transiently transfected cells showed that RNA molecules transcribed by CTD-truncated RNA Pol II were not efficiently polyadenylated (McCracken et al., 1997b). It was also shown that cleavage-polyadenylation factors CPSF and CstF present in unfractionated nuclear extracts specifically bound to CTD affinity columns and co-purified with RNA Pol II in a high-molecular-mass complex (McCracken et al., 1997b). Unlike capping and splicing factors, CPSF and CstF bind independently of the phosphorylation status of the CTD. Surprisingly, the recruitment of cleavage-polyadenylation factors to nascent transcripts occurs not only at the 3' end, but also at the promoter and throughout the length of the gene. ChIP experiments in yeast revealed that cleavage-polyadenylation factors are progressively recruited with a small but significant quantity at the 5' end which increases toward the 3' end (reviewed by Bentley, 2005). Accordingly, it has been shown that CPSF is brought to the pre-initiation complex by the general transcription factor TFIID and after transcription begins it dissociates

from TFIID and becomes associated with the elongating RNA Pol II (Dantonel et al., 1997). Furthermore, a biochemical study demonstrated that the CTD has a more direct role in 3' end processing than just the recruitment of the machinery, as it was shown that it is essential for the cleavage step of the polyadenylation reaction *in vitro* in the absence of transcription (Hirose and Manley, 1998; Ryan et al., 2002). These results suggested that the CTD participates directly in the formation and/or function of a stable, catalytically active processing complex through direct interaction with cleavage-polyadenylation factors.

As mentioned before, 3' end processing and termination of transcription are intimately coupled. This coupling ensures that RNA Pol II is only released from the template after it has completed synthesis of a full-length transcript. Two basic models of how these two processes are coupled were put forward (Figure 5). The first, initially called "antiterminator" model but now also know as the "allosteric" model, invokes that upon recognition of the poly(A) signal in the nascent transcript an allosteric change in the transcription complex decreases the processivity of the elongating polymerase making it termination competent. This could be due to the dissociation of an antiterminator from the elongating complex or association of a factor that reduces processivity (Figure 5A) (Buratowski, 2005; Logan et al., 1987). The other model called the "torpedo" model originally suggested that cleavage of the nascent transcript at the poly(A) site initiated termination by promoting rapid degradation of the 3' product still attached to the elongating polymerase. The 5' uncapped end of the transcript would be an entry point for a 5' to 3' exonuclease that would chase down the polymerase and induce termination (Figure 5B) (Connelly and Manley, 1988; Proudfoot, 1989). Although both models try to explain the coupling between 3' end processing and transcription termination, the "allosteric" model implies that termination is cleavage-independent whereas the "torpedo" model is based on the fact that termination is dependent on the cleavage reaction (reviewed by Rosonina et al., 2006).

Studies over the past years provided evidence for both models and indicated that they may not be mutually exclusive. The "allosteric" model of termination was supported by electron microscopy studies of nascent transcripts in *Xenopus* oocytes (Osheim et al., 1999) and *Drosophila* (Osheim et al., 2002), where terminated but uncleaved RNAs could be observed. This model also received support from work in yeast in which termination occurred in the absence of the cleavage reaction (Sadowski et al., 2003). It was recently found that a protein called Pcf11 is able to dismantle the elongating complex both in yeast (Zhang et al., 2005) and *Drosophila* (Zhang and Gilmour, 2006) in the absence of nucleotide hydrolysis.

Furthermore, yeast Pcf11 associates directly with both RNA and the CTD (Zhang et al., 2005). It is therefore possible that conformational changes in the CTD could be transduced through Pcf11 to the nascent transcript to cause termination in response to transcription of the poly(A) signal. On the other end, the "torpedo" model received support from the observation that mutants defective in 3' end cleavage were also defective in termination, providing evidence for a mechanism dependent on cleavage (Birse et al., 1998). Strong evidence for this model came also from the recent identification of the yeast Rat1 (Kim et al., 2004b) and the human Xrn2 (Gromak et al., 2006; West et al., 2004), both 5' to 3' RNA exonucleases necessary for efficient termination. Rat1 mutations strongly stabilize the RNA fragment downstream from the cleavage site and RNA Pol II fails to stop when it should (Kim et al., 2004b). Similarly, depletion of Xrn2 from human cells using RNA interference (RNAi) impairs termination on a transiently transfected β-globin gene (West et al., 2004). A recent report showed that in yeast, nascent RNA is in fact co-transcriptionally degraded downstream of the poly(A) site and the exonuclease Rat1 contribute to this co-transcriptional degradation (Luo et al., 2006). Nevertheless, this degradation is not sufficient to cause polymerase release because it can occur without eliciting termination. Surprisingly, Rat1 was found to functions in both 3' end processing and termination by enhancing recruitment of 3' end processing factors, including Pcf11 and Rna15. In addition, the cleavage factor Pcf11 reciprocally aids in recruitment of Rat1 to the elongation complex (Luo et al., 2006).

Analyses of the human β-globin gene indicated that, in addition to the normal cleavage/polyadenylation site, another cleavage site located downstream of the poly(A) site is important for transcription termination (Dye and Proudfoot, 2001). This site is co-transcriptionally cleaved (Dye and Proudfoot, 2001) and therefore was named co-transcriptional cleavage (CoTC) site (Figure 5C). It was recently shown that the RNA sequence at the CoTC forms a structure that has intrinsic self-cleavage activity in the absence of proteins, which is called an RNA enzyme or ribozyme (Teixeira et al., 2004). Mutations that inhibit self cleavage at the CoTC site also inhibit transcriptional termination. So far the CoTC element has only been described in primate β-globin genes (Teixeira et al., 2004), but it is possible that similar ribozymes may be present downstream of many other genes. In the presence of CoTC this is the entry site for the human 5' to 3' exonuclease Xrn2 (West et al., 2004), whereas in its absence the entry site can still be the cleavage/polyadenylation site (Gromak et al., 2006). This is what happens in the case of the human C2 complement gene and the human β-actin gene that instead of a CoTC site possess an RNA Pol II pause site

(Gromak et al., 2006). The 3' to 5' exonucleolitic activity that degrades the upstream cleavage product in the presence of CoTC remains to be determined but one proposed possibility is a complex of 3' to 5' exonucleases called the exosome (Figure 5C) (West et al., 2006).

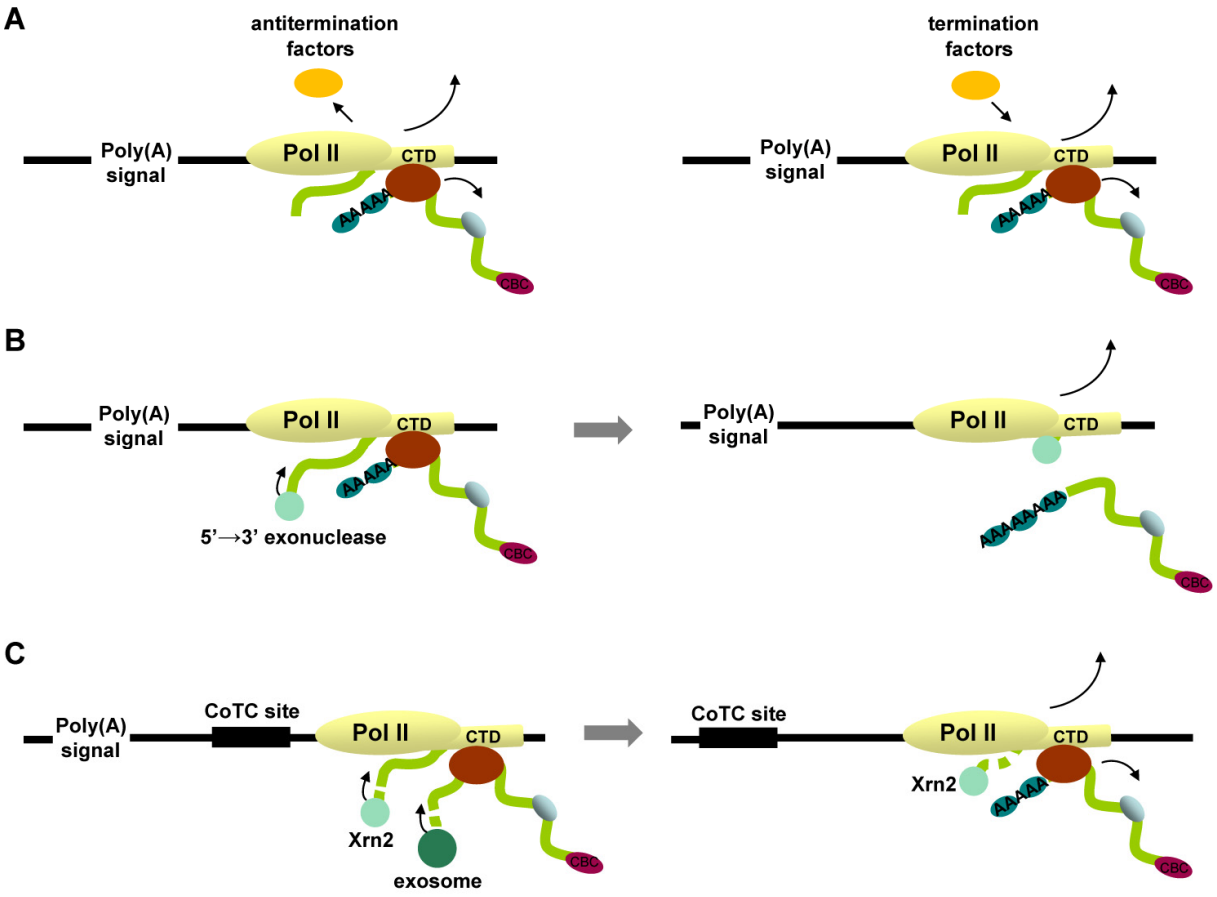


Figure 5 - Models for transcription termination by RNA Pol II. The transcripts are represented by a green line, the cleavage and polyadenylation factors are represented by a red oval and proteins that are bound to the mRNA forming a messenger ribonucleoprotein particle (mRNP) are blue ovals. **(A)** The allosteric model. Transcription termination is caused by the destabilization and/or a conformational change of the RNA Pol II elongation complex after transcribing the poly(A) site. Release of antitermination factors (*left*) or recruitment of termination factors (*right*) triggers dissociation from template DNA. **(B)** The torpedo model. Endonucleolytic cleavage at the poly(A) site creates an entry site for the 5' → 3' exonuclease (Rat1 in yeast and Xrn2 in humans), which degrades the RNA downstream of the cleavage site. Short nascent RNA may induce arrest of RNA Pol II elongation complex and promote termination. **(C)** The torpedo model in the presence of a co-transcriptional cleavage (CoTC) site. Autocatalytic CoTC acts as a precursor to termination by presenting a free 5' end that is recognised by the 5' → 3' exonuclease Xrn2. Subsequent degradation of this downstream cleavage product by Xrn2 leads to RNA Pol II transcriptional termination analogous to the original torpedo model shown in B. The 3' end of the cleaved RNA is likely to be degraded by the exosome. Adapted from (Rosonina et al., 2006; West et al., 2004).

The importance of the Rat1 and Xrn2 5' to 3' exonucleases in termination of transcription indicates that a conformation change in the elongation complex, as proposed by the "allosteric" model, is insufficient to promote termination, and strongly supports the "torpedo" model. On the other hand, the dependence of efficient termination on the cleavage/polyadenylation site in addition to the CoTC (West et al., 2004), plus the recent evidence indicating that exonuclease digestion of the downstream RNA is not sufficient to trigger termination (Luo et al., 2006), argue against the "torpedo" model. One possibility that reconciles both models is that the RNA Pol II complex with which the 3' end processing factors associated through the CTD achieves cleavage at the poly(A) site, degradation of the nascent downstream RNA, and undergoes allosteric changes that promote its release from the template (Rosonina et al., 2006; Tollervey, 2004). This is consistent with the fact that the efficiency of mRNA 3' end processing appears to be unaffected by deletion of the CoTC element and, conversely, CoTC occurs in the absence of upstream 3' end processing, but polymerase release from the DNA template requires both a functional poly(A) signal and CoTC (Dye and Proudfoot, 2001). The recent finding that the yeast Rat1 is not a dedicated termination factor, but is an integrated component of the cleavage/polyadenylation apparatus are also in agreement with this unified model for transcription termination (Luo et al., 2006).

In addition to these biochemical links between factors recruited to the 5' and 3' ends of the gene, a physical connection between the promoter and poly(A) site has been reported in yeast (Ansari and Hampsey, 2005; O'Sullivan et al., 2004). These studies point to the existence of gene loops, in which the termination region of a transcriptionally active gene can be physically linked to its promoter site. In this situation, termination can additionally serve to facilitate transcriptional reinitiation by the same RNA Pol II.

3.2 Interaction between processing steps

There is evidence that the 5' cap bound by cap binding complex (CBC) is implicated in interactions with splicing components at the adjacent 5' splice site for removal of the first intron (reviewed by Proudfoot et al., 2002). The mammalian CBC is required for efficient recognition of the cap-proximal 5' splice site by U1 snRNP during early spliceosome formation (Izaurralde et al., 1994; Lewis et al., 1996). Additional research suggested that the CBC positively influences the U6 snRNA 5' splice site interaction, possibly by affecting the displacement of U1 snRNP by U6 (O'Mullane and Eperon, 1998). In the yeast *S. cerevisiae*

conditional inactivation of Ceg1 resulted in an increase in the amount of unspliced pre-mRNA *in vivo*, demonstrating that these interactions are conserved in evolution (Fresco and Buratowski, 1996; Schwer and Shuman, 1996). Recent experiments in yeast show that the CBC is necessary, but not sufficient, for co-transcriptional spliceosome assembly (Gornemann et al., 2005).

Several reports have demonstrated that 3' end processing is coupled to the splicing of the terminal intron of a gene (reviewed by Proudfoot et al., 2002). This coupling was first reported in an *in vitro* study where it was shown that polyadenylation was stimulated by the upstream intron and this stimulation required a 3' splice site but not a 5' splice site (Niwa et al., 1990). Reciprocally, it was shown that mutation of the AAUAAA polyadenylation consensus sequence inhibited *in vitro* splicing of an upstream intron but not the removal of introns further upstream (Cooke et al., 1999; Niwa and Berget, 1991). These early observations of the dependence of splicing on polyadenylation signals and *vice versa* suggested the existence of interaction between polyadenylation and splicing factors.

In vivo studies indicated that the second and last intron of the human β-globin pre-mRNA was also necessary for correct and efficient 3' end formation (Collis et al., 1990). Further analysis indicated that efficient 3' end formation of the human β-globin transcripts is strongly enhanced by the presence of the terminal 60 nucleotides of the last intron (Antoniou et al., 1998). Specifically, mutations in the intronic polypyrimidine (Py) tract reduce 3' end processing efficiency *in vitro* and *in vivo* (Cooke and Alwine, 2002; Cooke et al., 1999; Millevoi et al., 2002). The Py tract is recognised by $U2AF^{65}$ at an early step in the assembly of the spliceosome and in addition to its crucial role in splicing, $U2AF^{65}$ binding to the Py tract of the last intron of the human β-globin pre-mRNA promotes 3' end processing (Millevoi et al., 2002). Additional studies on the mechanisms by which $U2AF^{65}$ stimulates pre-mRNA 3' end processing led to the identification of an interaction between $U2AF^{65}$ and a subunit of the human cleavage factor I (CF I_m) (Millevoi et al., 2006). Tethered functional analysis showed that the $U2AF^{65}$/CF I_m interaction stimulates *in vitro* 3' end cleavage and polyadenylation. There is also evidence that the 3' end processing machinery is able to stimulate splicing. It was shown that PAP can stimulate splicing through its interaction with $U2AF^{65}$ and as a consequence of this interaction there is increased binding of $U2AF^{65}$ to the Py tract of the 3' splice site adjacent to the 3' end formation signals (Vagner et al., 2000). Taken together these results suggest a major role for $U2AF^{65}$ in the coupling between splicing of the last intron and 3' end processing. In conclusion, $U2AF^{65}$ has been implicated in both

the stimulation of splicing by the 3' end processing machinery through it's binding to PAP (Vagner et al., 2000) and in the stimulation of 3' end processing by the splicing machinery through its binding to CF I_m (Millevoi et al., 2006).

3.3 Coupling transcription and processing with mRNA export

Splicing and mRNA export

The first evidence suggesting that splicing can enhance the efficiency of mRNA export came from the demonstration that pre-mRNA molecules containing a single intron are exported more efficiently from microinjected Xenopus oocyte nuclei than the same RNA molecules lacking the intron (Luo and Reed, 1999). Furthermore, an *in vitro* spliced mRNA and a cDNA version of the same RNA, though identical in sequence, were assembled into different mRNPs and only the spliced mRNP was exported rapidly and efficiently from injected oocyte nuclei (Luo and Reed, 1999). Based on these experiments the authors concluded that intron removal appeared to lead to the deposition of specific proteins that targeted the spliced mRNA for efficient nuclear export. The direct evidence that pre-mRNA splicing alters mRNP protein composition came one year later with the use of a novel *in vitro* cross-linking approach which allowed the identification of several proteins that associated with the mRNA exon-exon junctions as a consequence of splicing (Le Hir et al., 2000b). Further studies showed that the spliceosome stably deposits several proteins on the mRNA, probably as a single complex, that protects 8 nucleotides of the mRNA at a conserved position 20-24 nucleotides upstream of exon-exon junctions, in a sequence-independent manner (Le Hir et al., 2000a). Initial immunoprecipitation experiments revealed that this complex, called exon junction complex (EJC), was composed by the splicing-associated factors SRm160, DEK and RNPS1, UAP56, the mRNA-associated shuttling protein Y14 and the mRNA export factor REF1/Aly (Kataoka et al., 2000; Le Hir et al., 2001; Le Hir et al., 2000a; Luo et al., 2001). In the following years several other proteins were found to be part of this complex, like the Y14 partner Magoh (Kataoka et al., 2001), the nonsense-mediated mRNA decay (NMD) factors Upf3a/b, Upf2 and Upf1 (Le Hir et al., 2001; Lykke-Andersen et al., 2001), the nuclear translation-like factor eIF4AIII, a member of the DExH/D-box family of NTP-dependent RNA binding proteins (Ferraiuolo et al., 2004; Shibuya et al., 2004), the nucleocytoplasmic shuttling protein MLN51 (also known as Barentsz, Btz) which is

overexpressed in breast cancer (Degot et al., 2004), the RNP1 interacting protein Pinin (Li et al., 2003), SAP18, a shuttling protein, and Acinus, a protein restricted to the nucleus (Tange et al., 2005). The EJC is a highly dynamic structure consisting of a few core proteins plus several more peripherally associated factors (Figure 6). The core of the EJC is composed of four proteins: eIF4AIII, Y14, Magoh and MLN51. These proteins form a stable heterotetramer that remains bound to the mRNA throughout many different cellular environments (Tange et al., 2005), with eIF4AIII providing direct contact to the mRNA (Shibuya et al., 2004). All factors in this core are shuttling proteins and most likely follow the mRNA to the cytoplasm (Tange et al., 2005). Proteins in the outer shell were all found by mass spectrometry of an *in vitro*-derived EJC (Tange et al., 2005). RNPS1, Acinus, and SAP18 can stably associate (Schwerk et al., 2003) and may bind the EJC core as a trimeric complex. However, RNPS1 may also bind alone, via interactions with Pinin (Sakashita et al., 2004). SAP18, RNPS1, and REF1/Aly are shuttling proteins (Lykke-Andersen et al., 2001; Tange et al., 2005; Zhou et al., 2000), whereas Acinus and Pinin are nuclear restricted (Li et al., 2003; Tange et al., 2005). Transiently interacting factors are proteins not identified in an *in vitro*-derived EJC, but which likely interact dynamically with either the EJC core or outer sphere proteins (Tange et al., 2005).

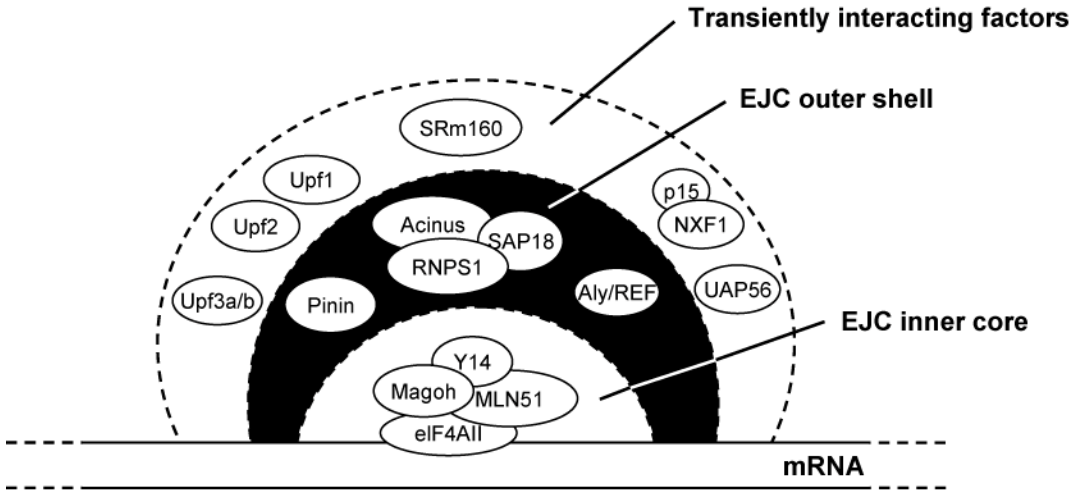


Figure 6 - Three spheres of exon junction complex (EJC) factors. The minimal EJC core likely consists of a tetrameric complex containing eIF4AIII, MLN51, Magoh, and Y14 (Tange et al., 2005). Proteins in the outer shell were all found by mass spectrometry of the *in vitro*-derived EJC (Tange et al., 2005). Transiently interacting factors are proteins not identified in the *in vitro*-derived EJC, but are likely to interact dynamically with either the EJC core or the outer shell proteins (see text for more details). Adapted from (Tange et al., 2005).

Despite being the focus of intense research, the assembly pathway of the EJC is still poorly understood. In human cells, the EJC is assembled during the second step of splicing, when the lariat has been formed and the exons are being ligated (Reichert et al., 2002; Shibuya et al., 2004). It was reported that numerous proteins interact with the 5' exon in the catalytically active spliceosome and during exon ligation, a major restructuring of this region occurs, placing three other proteins (p170, p95 and p57) at the site where the EJC forms (Reichert et al., 2002). At least one EJC component, REF1/Aly, can interact with the pre-mRNA prior to spliceosome assembly, whereas Y14, Magoh, RNPS1, UAP56, and SRm160 were found in intermediate-containing spliceosomes. Upon exon ligation, association of RNPS1, UAP56, and SRm160 is destabilized. In contrast, REF1/Aly, Y14, and Magoh remain stably bound to spliced mRNA (Reichert et al., 2002). Once formed, the EJC is associated with mRNA molecules that are bound by the nuclear cap-binding protein CBP80, both in the nucleus and in the cytoplasm, but not with mRNA molecules bound by the cytoplasmic eIF4E cap-binding protein, suggesting that at least part of the EJC is exported and then dissociates from the mRNA in the cytoplasm (Lejeune et al., 2002).

The nature of the proteins identified in the EJC led to the proposal of several functions for the EJC at different steps of mRNA metabolism, including nucleocytoplasmic transport of spliced mRNA, NMD (the degradation of mRNAs containing premature stop codons), mRNA localisation in the cytoplasm and translational yield of mRNA (reviewed by Tange et al., 2004). The complex is believed to be eventually removed by ribosomes in the first round of translation (Lejeune et al., 2002).

The notion that EJC deposition leads to recruitment of the mRNP export factor NXF1 (TAP) is an attractive model to explain the stimulatory effect of splicing on export. Evidence for this model came from an initial study showing that the EJC provides a binding platform for the mRNA export factors REF1/Aly and NXF1-NXT1 (also known as TAP-p15) (Le Hir et al., 2001). It was also shown in metazoan cells that REF1/Aly is specifically recruited to mRNPs generated by splicing (Zhou et al., 2000) and interacts with the spliceosomal protein UAP56 (Luo et al., 2001). Furthermore, recruitment of REF1/Aly to spliced mRNAs depends on the interaction with UAP56 (Luo et al., 2001) and REF1/Aly also interacts directly with NXF1 (Rodrigues et al., 2001; Stutz et al., 2000; Zhou et al., 2000). Therefore, a possible explanation for the splicing dependent recruitment of NXF1-NXT1 in metazoan cells is that during splicing UAP56 facilitates spliceosome assembly and then recruits REF1/Aly that in turn recruits the NXF1-NXT1 heterodimer, which then targets the spliced mRNPs to the NPC

(reviewed by Cullen, 2003). In metazoans, where most genes contain introns, it makes perfect sense that the conserved machinery for mRNA export is intimately tied to splicing. But could this also occur in *S. cerevisiae,* where only a few genes contain introns? Experimental evidence indicates that it is possible that both splicing dependent recruitment of the mRNA export machinery and EJC formation are conserved in yeast. In *S. cerevisiae*, the counterparts of REF1/Aly (Yra1p) and UAP56 (Sub2p) also interact with each other directly and are required for mRNA export (Strasser and Hurt, 2001) and Yra1p also interacts directly with the yeast ortholog of metazoan NXF1 (Mex67p) (Strasser and Hurt, 2000).

A new set of experiments using an RNA interference (RNAi) approach both in *Drosophila* cells and in *C. elegans* brought some confusion to the emerging picture of a conserved mRNP export pathway linked to the EJC (the UAP/REF/NXF1 pathway). The conclusions of the RNAi study in *Drosophila* were that although NXF1, NXT1 and UAP56 were required for mRNA export (Gatfield and Izaurralde, 2002; Herold et al., 2001; Herold et al., 2003), REFs and the other EJC components were dispensable (Gatfield and Izaurralde, 2002). The conclusions of the *C. elegans* study were similar with NXT1 being required for mRNP export but the REF proteins (there are three REF proteins present in *C. elegant*) and several EJC proteins being dispensable (Longman et al., 2003). There are several possibilities to explain these results: either the RNAi-mediated knockdown of REF and the other EJC proteins was not sufficient to eliminate their functions or there may be a redundancy in factors (Reed, 2003). Another possibility is that the splicing dependent recruitment of the export factors to the mRNPs is not essential for export but it increases the efficiency and/or fidelity of the process by promoting a more efficient recruitment (Ohno et al., 2002). The authors of the above studies favour yet another possibility. They propose that other adaptor protein(s) may mediate the interaction between NXF1 and the mRNA. In agreement with this hypothesis, members of the SR family of splicing factors have been proposed to fulfil this role (reviewed by Reed and Cheng, 2005). In metazoans, these proteins bind to exon sequences in pre-mRNA and recruit the spliceosome to the flanking 5' and 3' splicing sites. After splicing, SR proteins remain bound to the spliced mRNA (see Graveley, 2000) and shuttle between the nucleus and cytoplasm (Caceres et al., 1998) as would be expected of an mRNA export protein. In fact, it was shown that the SR proteins 9G8 and SRP20 mediate mRNA export (Huang and Steitz, 2001) and function by direct interaction with NXF1 (Huang et al., 2003). Furthermore, it was shown that phosphorylation of the SR proteins is involved in regulating their interaction with NXF1 (Huang et al., 2004). Specifically, the dephosphorylated form of

SR proteins associates with mRNA and with NXF1 and thus this dephosphorylation is thought to be a mechanism for the selective export of spliced mRNA versus unspliced pre-mRNA (Huang et al., 2004).

Transcription and mRNA export

Although the UAP/REF/NXF1 export machinery seems to be conserved from yeast to humans, it cannot be exclusively recruited by a splicing-dependent mechanism. In fact, 95% of *S. cerevisiae* genes as well as several metazoan genes lack introns. Accordingly, it was shown that the *S. cerevisiae* Sub2p and Yra1p (human UAP56 and REF1/Aly, respectively) are essential for nuclear export of the intronless mRNPs (Jensen et al., 2001a; Strasser and Hurt, 2000; Strasser and Hurt, 2001) and Yra1p binding to mRNPs can occur in the absence of splicing (Lei and Silver, 2002a). In metazoan cells, recruitment of UAP56 to mRNPs has also been shown to occur independently of splicing (Kiesler et al., 2002) and knockdown of either UAP56 or NXF1 by RNAi blocks the nuclear export of both spliced and intronless transcripts (Gatfield et al., 2001; Herold et al., 2001). Thus, a central focus of recent research has been to determine the mechanisms for splicing-independent recruitment of this machinery. One hypothesis that has been proposed is that transcription itself can promote the recruitment of nuclear export factors.

The first evidence for this hypothesis came from ChIP experiments in *S. cerevisiae* showing that the mRNA export machinery is recruited co-transcriptionally (Lei et al., 2001). New evidence came from the finding that Sub2p and Yra1p are stoichiometrically associated with the yeast THO complex (Strasser et al., 2002). This complex was previously identified as a four protein complex, composed of stoichiometric amounts of Tho2, Hpr1, Mft1 and Thp2 and shown to be involved in transcription elongation and the maintenance of genome stability (Chávez et al., 2000). The THO complex together with the mRNA export proteins was designated “transcription/export” (TREX) complex (Strasser et al., 2002). It was also shown that Sub2 and Yra1 interact genetically with all four components of the THO complex and that these components operate in the export of bulk poly(A)$^+$ RNA as well as of mRNA derived from intronless genes. Both REF1/Aly and UAP56 associate with human counterparts of the THO complex suggesting that the TREX complex has a conserved role in coupling transcription to mRNA export (Strasser et al., 2002). Further studies showed that the THO component Hpr1 interacts directly with Sub2p and plays a role in recruiting Sub2p and Yra1p to actively transcribed genes (Zenklusen et al., 2002). In a genetic screen in *S. cerevisiae,* an

interaction between Yra1p and the conserved protein Sac3p was identified (Fischer et al., 2002). Sac3p was previously found to interact with Sub2p (Strasser et al., 2002), form a stable complex *in vivo* with Thp1p, a protein previously reported to function in transcription elongation (Fischer et al., 2002) and bind to the mRNA export factor Mex67p-Mtr2p (human NXF1-NXT1) (Fischer et al., 2002; Lei et al., 2003). Moreover, Sac3p interacts with the nucleoporin Nup1p located at the nuclear side of the nuclear pore complex (NPC) (Fischer et al., 2002). Immunoelectron microscopy analysis showed that Sac3p localises exclusively to cytoplasmic fibrils of the NPC and Mex67 accumulates at the nuclear rim when SAC3 is mutated (Lei et al., 2003). Significantly, mutations in Sac3p or Thp1p lead to strong mRNA export defects (Fischer et al., 2002; Lei et al., 2003). On the basis of these observations, a model can be proposed in which Sac3p first associates with the mRNP early in its formation, perhaps during transcription, and then by interacting with both Mex67p and the nucleoporin, facilitates docking of the mRNP at the NPC and functions in the translocation of Mex67p across the NPC (Fischer et al., 2002; Lei et al., 2003). Metazoan counterparts of Sac3 and Thp1 have been identified. Thus, it would not be surprising if these proteins played also a role in mRNA export in metazoans.

ChIP assays revealed that THO complex components, as well as Yra1p and Sub2p, are recruited to actively transcribed genes during transcription and travel the entire length of the gene with RNA Pol II (Abruzzi et al., 2004; Strasser et al., 2002; Zenklusen et al., 2002). A modified ChIP assay that includes an RNase step indicates that Sub2p binds to nascent RNA, Yra1p associates with both RNA and DNA, and Hpr1p associates only with DNA (Abruzzi et al., 2004). Although Hpr1p was found to be recruited similarly to both intronless and intron-containing genes, low Yra1p and Sub2p levels were present on a subset of intron-containing genes. The residual Yra1p and Sub2p recruitment was less RNA-associated and showed a good correlation with high levels of U1 snRNP on these genes. These experiments support a model in which TREX is recruited via the transcription machinery and then Yra1p and Sub2p are transferred to the nascent RNA (Figure 7A). On some intron-containing genes, retention and/or transfer of Yra1p and Sub2p to nascent RNA seems to be inhibited (Abruzzi et al., 2004).

The *Drosophila* counterpart of the THO complex was identified and contains homologs of Tho2 and Hpr1 as well as three proteins, THOC5, THOC6 and THOC7, which do not have apparent yeast homologs (Rehwinkel et al., 2004). RNAi showed that the *Drosophila* THO complex functions in mRNA export but genome-wide studies led to the

conclusion that the vast majority of mRNAs are transcribed and exported independently of the THO complex (Rehwinkel et al., 2004).

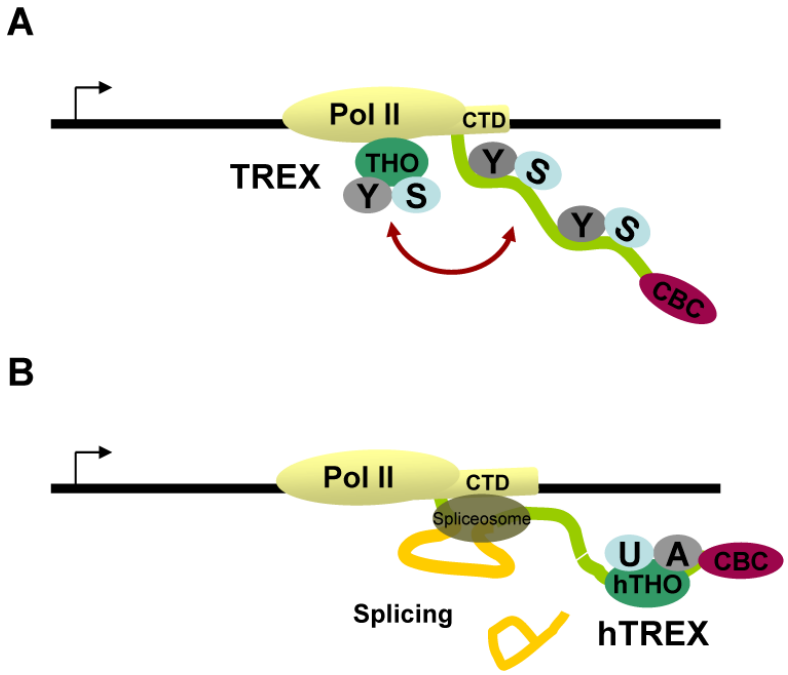


Figure 7 - Models for "transcription/export" (TREX) complex recruitment. (A) Co-transcriptional recruitment of the TREX complex in yeast. mRNA export proteins Yra1 (Y), Sub2 (S) associate with the THO complex to form the TREX complex. This complex is recruited to active genes and functions to load Yra1 and Sub2 onto the nascent transcript. **(B)** Recruitment of the human TREX complex during splicing. The hTREX complex is composed by the hTHO complex, UAF56 (U) and REF1/Aly (A). Spliceosome assembly and splicing occur as the transcript is synthesized by RNA Pol II and the hTREX complex is recruited to the spliced mRNA near the 5' end, with REF1/Aly bound closest to the cap binding complex (CBC). For simplicity, other components of the spliced mRNP, such as EJC proteins, are not shown. Adapted from (Masuda et al., 2005; Reed and Cheng, 2005).

Recently, the human TREX complex was characterised in detail and shown to contain REF1/Aly, UAP56 and the human counterpart of the THO complex (Masuda et al., 2005). Surprisingly, *in vitro* studies indicate that the human THO complex associates with spliced mRNA but not with unspliced pre-mRNA and this association is independent of transcription (Masuda et al., 2005). These new data suggests that in humans, recruitment of the TREX complex to spliced mRNA is not directly coupled to transcription as it is in yeast, but is instead coupled to splicing and only indirectly to transcription because splicing in humans is a co-transcriptionally event (Masuda et al., 2005). Assuming that this splicing-coupled mechanism also occurs *in vivo* in human cells, one hypothesis that could reconcile these

results with what was shown in yeast is that there are two mechanisms for TREX complex recruitment: a splicing-coupled more prevalent in metazoans, and a direct transcription coupled, more common in yeast (Figure 7). These mechanisms may differ in yeast *versus* metazoans because most yeast genes lack introns (for review see Reed and Cheng, 2005). Recent data (Cheng et al., 2006) helped to understand better the splicing-coupled mechanism of TREX recruitment in human cells. This study revealed that the human TREX complex is recruited in a splicing- and cap-dependent manner to a region near the 5' end of the mRNA with the TREX component REF1/Aly bound closest to the 5' cap (Figure 7B). The CBP80, which is bound to the cap, associates efficiently with TREX, and REF1/Aly mediates this interaction suggesting that CBP80-REF1/Aly interaction results in recruitment of TREX to the 5' end of mRNA, where it functions in mRNA export (Cheng et al., 2006).

3' end processing and mRNA export

Several studies have demonstrated a connection between mRNA 3' end processing and export. In human cells, mRNA molecules bearing 3' ends formed by a self-cleaving ribozyme element were only detected in the nucleus, indicating that they were not efficiently exported to the cytoplasm (Eckner et al., 1991; Huang and Carmichael, 1996). Indeed, the proper completion of 3' end formation is required for efficient nuclear export of mammalian mRNAs and it appears that it is the process of polyadenylation, and not merely the presence of a poly(A) tail, that is crucial for the generation of an export-competent mRNA (Eckner et al., 1991; Huang and Carmichael, 1996).

In *S. cerevisiae* two observations initially suggested that the transport of the mRNA from the nucleus to the cytoplasm is also sensitive to proper 3' end formation, including polyadenylation. First, mRNAs lacking a proper 3' end formation signal accumulated within the nucleus (Long et al., 1995). Second, fluorescent labelling of specific transcripts in living yeast cells demonstrated their nuclear accumulation when assayed in strains defective in 3' end formation (Brodsky and Silver, 2000). In an effort to understand mRNA export and its relationship to other aspects of mRNA metabolism, numerous mutants that affect mRNA export in yeast were isolated. These mRNA export mutants were identified by several criteria, including screenings for lesions that result in the nuclear accumulation of polyadenylated RNA. These screenings led to the identification of genes important for ribonucleic acid trafficking (RAT), mRNA transport (MTR) and mRNA export (MEX) (see Hilleren and Parker, 2001). Yeast strains carrying lesions in several of these genes produce nascent

transcripts carrying poly(A) tails roughly 30 residues longer than the nascent poly(A) tails observed in wild-type strains (Hilleren and Parker, 2001). Furthermore, several factors involved in termination and 3' end processing were also obtained in a screening to identify genes required for mRNA export in yeast (Hammell et al., 2002). Taken together these observations provide support for the view that, at least in yeast, mRNA 3' end formation and mRNA export are mechanistically coupled events.

Although substantial evidence indicate that in yeast the mRNA export machinery is co-transcriptionally loaded onto mRNA, 3' end formation may play a more critical role in this loading. For example, Sub2p is not required for the recruitment of Yra1p to genes that lack introns. However, proper 3' end formation is required to recruit this factor, whether or not an intron is present (Lei and Silver, 2002a). In further support of this view it was reported that transcripts generated by T7 polymerase (i.e. not by RNA Pol II) are exported only if 3' end formation occurs normally (Dower and Rosbash, 2002).

Although all these data clearly indicate some form of tight mechanistic coupling between mRNA 3' end formation and nuclear export, its molecular basis remains unresolved.

4. Integrating gene expression with nuclear architecture

Most of the knowledge collected on each step of gene expression came from biochemical experiments, where they were studied individually. Discoveries within the last decade have revealed that the multiple steps in gene expression are remarkably coupled and this coupling must reflect their spatial organisation inside the cell. For this reason, a better understanding of gene expression must take into account its natural cellular context and in particular the nuclear architecture. Cell biological approaches are essential to study gene expression *in vivo*, since they allow analysis of gene expression at the single cell level, in living cells and in a spatial and temporal context.

4.1 The concept of sub-nuclear compartments

It is now accepted that the nucleus is a highly dynamic and organised organelle, functionally and spatially compartmentalised (reviewed by Misteli, 2000; Misteli, 2005;

Misteli, 2008). The interphase nucleus can be divided into chromatin containing regions and interchromatin nucleoplasmic space. In the chromatin regions each chromosome occupies a nuclear space or territory and specific DNA sequences within each chromosome are organised as either euchromatin (less condensed genome regions) or heterochromatin (highly condensed regions) (Meaburn and Misteli, 2007). The classic view correlates chromatin morphology with gene activity, where euchromatin is considered to be transcriptionally active and heterochromatin considered transcriptionally silenced (reviewed by Cremer and Cremer, 2001). A recent study, however, found a better correlation between chromatin morphology and gene density, with gene-rich regions being present in open chromatin regions and gene-poor regions in more condensed chromatin (Gilbert et al., 2004).

In the interchromatin space several sub-nuclear compartments may form. The best characterised is the nucleolus, the site of rRNA synthesis. In addition, several nuclear bodies, often referred as foci, because of their appearance by fluorescence microscopy, are present in the nucleus. The best characterised nuclear bodies are the Cajal body (involved in snRNP biogenesis) and the PML body (promyelocytic leukaemia body, riche in transcriptional regulators) (reviewed by Zhong et al., 2000). Additional compartments that occupy a large volume of the interchromatin nucleoplasmic space are the splicing factor compartments (SFC) also called speckles because of their irregular shape when detected by immunofluorescence using anti-splicing factor antibodies. By electron microscopy the SFC correspond to two distinct structures that cannot be distinguished by fluorescence microscopy, the interchromatin granule clusters (IGC) and the perichromatin fibrils (Spector, 1993). The IGC are composed of clusters of 20-25 nm in diameter that seem to be connected by a thin fibril. The perichromatin fibrils are fibrillar structures of 3-5 nm in diameter found at the periphery of IGC. Several lines of evidence point to IGC as storage/ assembly/ modification compartments that can supply splicing factors to active transcription sites (reviewed by Lamond and Spector, 2003). Live-cell studies show that splicing factors are recruited from IGC to sites of transcription (Misteli et al., 1997). When transcription or pre-mRNA splicing are inhibited, splicing factors accumulate in enlarged, rounded IGC (Spector et al., 1983) (O'Keefe et al., 1994). Thus, the IGC are dynamic structures that change in response to the transcriptional activity of the cell.

4.2 Organisation of RNA polymerase II transcription sites in the nucleus

The first experiments that allowed the visualisation of transcriptional activity in the mammalian nucleus consisted of short pulses of tritiated uridine that were analysed at the electron microscopic level. These experiments showed that nascent pre-mRNA is predominantly localised in the perichromatin fibrils at the periphery of IGCs and in other regions throughout the nucleoplasm (reviewed by Fakan, 1994). More recent results using incorporation of bromouridine triphosphate (Br-UTP) that can be detected by immunolabelling and visualised at the fluorescence or electron microscope allowed a more detailed analysis of transcription sites. These experiments showed that the transcription sites appear as discrete foci throughout the nucleoplasm and confirm their association with perichromatin fibrils. RNA Pol II itself is also distributed widely throughout the nucleoplasm in many discrete sites that frequently coincides with the foci of labelled nascent transcripts (reviewed by Szentirmay and Sawadogo, 2000). The number of RNA Pol II transcription sites in HeLa cells has been estimated using different RNA labelling and cell preparation methods and combining fluorescence and electron microscopy. These experiments provided an estimate of several thousand transcription foci for RNA Pol II in each HeLa cell nucleus (see Szentirmay and Sawadogo, 2000). An estimate of the number of active polymerases determined that each focus may contain 8-15 transcription polymerases giving rise to the proposal that RNA Pol II associates into “transcription factories” (reviewed by Cook, 1999; Pombo et al., 2000). This could represent a way to increase the local concentration of transcription factors and RNA Pol II at sites of active transcription. The observation that active RNA Pol II transcription units are resistant to electroelution, even after the removal of most of the chromatin, supports a model in which the polymerases are attached in the nucleus (Iborra et al., 1996; Jackson et al., 1996). According to this model, the genes are recruited to and transcribed by attached polymerases. The loops of chromatin surround the transcription factories and the DNA moves past the attached polymerases during transcription (Iborra et al., 1996).

4.3 Pre-mRNA processing within the nucleus

Pre-mRNA processing was initially considered a post-transcriptional process but there is now definitive biochemical, cytological and genetic evidence that much of the pre-mRNA

processing takes place co-trancriptionally (see Szentirmay and Sawadogo, 2000). The first unambiguous demonstration that pre-mRNA splicing takes place co-trancriptionally was provided by electron micrographs of Miller chromatin spreads of *Drosophila* embryo genes, where ribonucleoprotein particle formation and intron removal could be seen on nascent transcripts (Osheim et al., 1988). In mammalian cells, Fakan *et al.* (Fakan et al., 1986) localised snRNPs to elongating RNP fibrils and in *Chironomus tentans* the spliceosomes were found to be associated with nascent RNA molecules transcribed from the Balbiani ring genes (Kiseleva et al., 1994).

There were for some time discrepancies in the literature concerning the possible direct role of SFC as splicing sites, but there is now substantial evidence that co-transcriptional splicing is associated with perichromatin fibrils, rather than, with IGC. These discrepancies might have arisen essentially for two reasons. Firstly, because it is not possible to distinguish perichromatin fibrils from IGC, using fluorescence microscopy. Secondly, because in these studies highly transcribed genes were used and they recruit a significant amount of pre-mRNA splicing factors, making these regions indistinguishable from IGC in the fluorescence microscope (Lamond and Spector, 2003).

Several experimental approaches indicate that, rather than being sites of active splicing, the SFC/IGC function mostly as storage compartments that can supply splicing factors to active transcription sites (see Misteli and Spector, 1998)). The first clues for the role of SFC/IGC in splicing came from work on adenovirus-infected cells. During intermediate stages of infection, viral transcripts accumulate in the interchromosomal space and it was observed that splicing factors also accumulate at these sites of viral infection, resulting in reduced SFC. These experiments let to the proposal that splicing factors are recruited from SFC to sites of active transcription upon transcriptional activation of genes (reviewed by Misteli and Spector, 1998). These results were later confirmed using transfected DNA templates encoding either intron-containing or intronless transcripts in which the splicing factors were only localised to the sites of intron-containing transcripts (Huang and Spector, 1996). Taking advantage of the green fluorescent protein (GFP), this recruitment model was directly tested in living cells by the visualisation of the dynamic properties of the pre-mRNA splicing factor SF2/ASF by time-lapse microscopy. These experiments showed that upon activation of genes, splicing factors moved from SFC to newly formed sites of transcription (Misteli et al., 1997). Based on the previously described experimental evidence, a model was proposed to explain the spatial organisation of pre-mRNA splicing in the mammalian cell

nucleus. In a transcriptionally active nucleus splicing factors localised in IGC represent storage sites while the splicing factors present in perichromatin fibrils are actively engaged in splicing. In response to transcriptional activity the splicing factors are recruited from IGC to perichromatin fibrils.

4.4 Dynamics of mRNA within the nuclear environment

After being released from the transcription sites, mRNA molecules must reach the NPC in order to be translocated to the cytoplasm for translation. How the mRNA molecules travel from the transcription site to the nuclear pore remains the subject of most investigation and is in general poorly understood (reviewed in Darzacq et al., 2005; Gorski et al., 2006; Misteli, 2008; Shav-Tal et al., 2006). As previously mentioned, during processing nascent mRNA molecules assemble together with RNA binding proteins into ribonucleoprotein particles (mRNPs). Because the intranuclear environment is very viscous, the simple diffusion of large macromolecular complexes such as mRNPs, was expected to be too slow to explain the fact that mRNAs could be detected in the cytoplasm within a few minutes after their synthesis. Therefore, the hypothesis that active genes are situated near the nuclear periphery and that mRNAs exit the nucleus through the nearest pores (known as the "gene gating" hypothesis) was proposed (Blobel, 1985). However, the development of new tools to study the movement of molecules in the nucleus such as photobleaching and fluorescence correlation spectroscopy techniques have indicated that macromolecule-sized solutes can diffuse in the nucleus at rates that are only approximately three- to fourfold slower than those observed in aqueous solution (Politz et al., 1998; Seksek et al., 1997). Reports addressing specifically the movement of nuclear mRNPs indicate that they move from the sites of transcription to the nuclear pores by a diffusion-based mechanism (Calapez et al., 2002; Molenaar et al., 2004; Politz et al., 1998; Politz et al., 1999; Shav-Tal et al., 2004; Snaar et al., 2002; Vargas et al., 2005), although energy dependent processes may also be involved (Calapez et al., 2002; Shav-Tal et al., 2004; Vargas et al., 2005). Diffusion coefficients from 0.01 to 0.6 $\mu m^2 s^{-1}$ have been estimated for nuclear mRNPs using labelling methods based either on fluorescent oligonucleotide probes (Molenaar et al., 2004; Politz et al., 1998; Politz et al., 1999; Vargas et al., 2005) or specific mRNA binding GFP-fused proteins (Calapez et al., 2002; Shav-Tal et al., 2004). Tracking of a single mRNP in the nucleoplasm of living mammalian cells indicated that it moves freely by Brownian diffusion (Shav-Tal et al., 2004;

Vargas et al., 2005). Furthermore, the diffusion of mRNPs is restricted to the extranucleolar, interchromatin spaces and if the mRNPs move into dense chromatin, they tend to become stalled. In this situation ATP is required for the complexes to resume their motion, but the movement of mRNPs does not require energy (Shav-Tal et al., 2004; Vargas et al., 2005).

5. Quality control mechanisms of gene expression in the nucleus

The accurate function of all the molecular machines in a cell depends on the integrity of their protein components which is directly related to both genome integrity and accuracy of all the steps of gene expression. Aberrant transcripts can arise from mutations in the genomic sequences or from errors generated during several steps of mRNA biogenesis. Recent studies revealed a complex network of interactions between components of the gene transcription, pre-mRNA processing, mRNA transport and mRNA translation machineries. Besides providing the means to regulate gene expression at many different levels, these interconnected processes provide also opportunities for quality control checkpoints, so that only fully processed and error-free mRNAs are translated into proteins (reviewed by Fasken and Corbett, 2005; Saguez et al., 2005; Sommer and Nehrbass, 2005).

The classical example of a quality control mechanism of gene expression is the recognition and degradation of transcripts harbouring premature termination codons, also known as nonsense-mediated decay (NMD). In fact, for a long time this was the only known mechanism of mRNA quality control or mRNA surveillance and is still the best well documented in several experimental systems. In recent years a number of mRNA quality control checkpoints which operate at different levels of mRNA biogenesis both in the nucleus and in the cytoplasm have been identified and this is currently a field of intense research (see Fasken and Corbett, 2005; Saguez et al., 2005; Sommer and Nehrbass, 2005). The first level of quality control of mRNA biogenesis must take place in the nucleus. In this cellular compartment transcripts are made and matured and the ultimate goal of the nuclear quality control checkpoints is that only perfect mRNA molecules cross the NPC. Two locations are emerging as “hot spots” for quality control in the nucleus, the site of transcription and the nuclear periphery. The molecular mechanisms of these checkpoints are still poorly understood and much more data has been collected in yeast than in mammalian systems.

5.1 A nuclear turnover pathway for mRNA

The first evidence for a quality control mechanism of gene expression operating in the nucleus came from experiments showing that pre-mRNA molecules bearing mutations that allow spliceosome assembly but impaired the splicing reaction were largely retained in this cellular compartment (Chang and Sharp, 1989; Legrain and Rosbash, 1989). Further experiments showed that mutated splicing reporter constructs were rapidly degraded in the nucleus of wild-type yeast cells (Burgess and Guthrie, 1993) and many yeast splicing mutants showed a strong reduction in mRNA levels without a corresponding accumulation of the unspliced precursor. These early observations suggested the existence of a nuclear turnover pathway to recognise and rapidly degrade unspliced pre-mRNA in yeast. The demonstration that such a pathway existed was finally reported in 2000 (Bousquet-Antonelli et al., 2000). This pathway involves a 3' to 5' degradation by a complex of exoribonucleases called the exosome and to a lesser extent also 5' to 3' degradation by the exoribonuclease Rat1p. The yeast exosome contains at least ten essential components, nine of which are 3' to 5' exoribonucleases and is active both in the nucleus and in the cytoplasm. The cytoplamic exosome participates in mRNA turnover. The nuclear exosome has additional components (Rrp6p, Rrp47p and Mtr4p) and was previously shown to be involved in the processing and degradation of small nuclear RNAs (snRNAs), small nucleolar RNAs (snoRNAs) and ribosomal RNAs (rRNAs) (Butler, 2002; Houseley et al., 2006). Mutation of nuclear exosome components in cells that are defective for splicing resulted in the nuclear accumulation of pre-mRNA, suggesting a direct competition between the splicing machinery and the nuclear pre-mRNA turnover pathway to degrade aberrant, unspliced transcripts (Bousquet-Antonelli et al., 2000). Another evidence for the quality control function of the nuclear exosome came from the observation, in yeast, that deletion of the essential *RRP6* (ribosomal RNA processing 6) gene (Briggs et al., 1998) could suppress a poly(A) polymerase mutant (pap1-1) revealing that Rrp6p was required for the degradation of transcripts with abnormal poly(A) tails (Burkard and Butler, 2000). It was also suggested that Rrp6p could be implicated in the decay of transcripts that do not acquire the correct complement of mRNA-binding proteins during their biogenesis (Zenklusen et al., 2002). Recent studies also showed that Rrp6p is required for the degradation of apparently normal mRNA that is retained in the nucleus through a pathway referred to as DRN (for decay of RNA in the nucleus). In addition to Rrp6p, the yeast DRN pathway depends on the 5' to 3' nuclear exoribonuclease Rat1p and on Cbc1p, a component

of the nuclear cap-binding complex (Das et al., 2003). It is not known whether the DRN pathway can also target abnormal transcripts and overlap with the exosome-mediated turnover pathway to degrade aberrant transcripts. It is conceivable to speculate that all transcripts that remain within the nucleus longer that they should are subject to a surveillance system and targeted for degradation (Kuai et al., 2005). The degradation could be mediated either by the DRN pathway if they are properly processed or by an exosome-mediated turnover pathway for aberrant transcripts (see Fasken and Corbett, 2005).

The mRNA turnover pathways operating in the nucleus are still poorly characterised and most of the available data came from yeast. All the exosome subunits are conserved in humans, where the complex is known as PM/Scl particle, a target of autoimmune antibodies, but it is still not known if the human exosome fulfils all the roles that have been attributed to the yeast exosome (reviewed by Schilders et al., 2006).

5.2 A polyadenylation-dependent checkpoint at the transcription site

Several observations provided support for the view that mRNA 3' end formation and mRNA export are mechanistically coupled events. Among these, one interesting observation was that yeast strains carrying lesions in several genes important for mRNA export produce nascent transcripts carrying poly(A) tails roughly 30 residues longer than the nascent poly(A) tails observed in wild-type cells (Hilleren and Parker, 2001). Moreover, analysis of the fate of the yeast heat shock transcripts *SSA4* and *HSP104* in a background of mRNA export mutants (rat7-1, mex67-5, and rat8-2 strains) confirmed the hyperadenylation phenotype and further showed that the specific transcripts were concentrated in a single intranuclear focus compatible with the site of transcription (Jensen et al., 2001b; Thomsen et al., 2003). This implies that polyadenylation and nuclear export are coupled in a step that involves the release of mRNA from the transcription site. Interestingly, unadenylated transcripts from the same genes, resulting from a pap1-1 mutant background, were also retained in the nucleus in discrete intranuclear foci compatible with transcription sites (Hilleren et al., 2001). The sequestration of both, unadenylated and hyperadenylated mRNAs, in transcription site foci indicated that a system exists in yeast to monitor the quality of 3' end formation and inhibit the release of aberrant transcripts. Because defects in the nuclear exosome component Rrp6p partially rescued the temperature sensitivity of pap1-1 (Burkard and Butler, 2000), the effect of an *RRP6* deletion on transcript localisation was investigated. In Rrp6p mutants both hypo-

as well as hyperadenylated mRNAs are released from the nuclear focus suggesting that Rrp6p and the nuclear exosome contributes to a checkpoint that monitors proper 3' end formation of mRNA (Hilleren et al., 2001). It was also found that nuclear *HSP104* transcripts produced in *sub2*, *hpr1* and other THO complex mutant strains are either 3' end truncated or retained at the transcription site in a Rrp6p-dependent manner (Libri et al., 2002). The localisation of the 3' end truncated transcripts is unclear and this phenotype can be reverted by deletion of *RRP6* which indicates that the 3' end truncation results from incomplete nuclear degradation rather than from a failure to efficiently elongate them (Libri et al., 2002). Furthermore, a plasmid-produced mRNA containing a self-cleaving hammerhead ribozyme element in place of cleavage and polyadenylation signals was shown to be predominantly unadenylated and exhibited a partial defect in export accumulating near its site of synthesis (Dower et al., 2004). In this case the nuclear accumulation of the transcripts was exosome-independent and could be relieved by insertion of a stretch of 48 or more DNA-encoded adenosine residues immediately upstream of the ribozyme element (a synthetic A tail). This indicates that, at least when the RNA is produced out of its normal chromatin context, a 3' stretch of adenosines can promote export, independently of cleavage and polyadenylation (Dower et al., 2004). A precise control of the poly(A) tail length seems to be important for mRNA release from sites of transcription. In yeast, three proteins were described to be important for the control of the poly(A) tail length *in vivo*, the poly(A) binding proteins Nab2p and Pab1p and the poly(A) nuclease, PAN (Dunn et al., 2005; Hector et al., 2002). Interestingly, loss of NAB2 expression leads to hyperadenylation and nuclear accumulation of poly(A)$^+$ RNA but, in contrast to mRNA export mutants, these defects can be uncoupled in a nab2 mutant strain (Hector et al., 2002). Pab1p contains a nonessential leucine-rich nuclear export signal and shuttles between the nucleus and the cytoplasm. It can exit the nucleus either as cargo of exportin 1 or bound to mRNA. Pab1p is essential but several bypass suppressors have been identified. Deletion of *PAB1* from these bypass suppressor strains results in exosome-dependent retention at sites of transcription (Dunn et al., 2005). Retention is also observed in cells lacking PAN. These nuclease complex is thought to be recruit by Pab1p and is responsible for trimming of the poly(A) tail to the length found on newly exported mRNAs (Dunn et al., 2005).

Recently a complex called TRAMP (*Tr*f4p/Air2p/*M*tr4p polyadenylation complex) has been identified as an exosome activating complex (LaCava et al., 2005). *In vitro*, the TRAMP complex showed distributive RNA polyadenylation activity and is believed to ease exosome

access onto difficult substrates (e.g. highly structured ones) by adding an unstructured tag. TRAMP-assisted exosomal decay might subsequently occur through multiple rounds of adenylation/degradation (LaCava et al., 2005). It was recently shown that TRAMP complexes harbouring Trf4p are involved in nuclear mRNA quality control in *S. cerevisiae* (Rougemaille et al., 2007). However, in contrast to Rrp6p, Trf4p is not required for retention in nuclear foci of either the inducible *HSP104* RNA or the constitutively expressed *PDR5* RNA. Transcription pulse–chase experiments showed that *HSP104* transcripts undergoing quality control in THO/sub2 mutant strains are partitioned into two different pools of molecules, one that undergoes rapid decay and another of stable RNAs retained in transcription site foci (Rougemaille et al., 2007).

It is now clear that there is an Rrp6p-dependent retention at the transcription site of heat shock mRNAs produced in the context of mRNA export factor mutants as well as polyadenylation-malfunctioning mutants and strains deleted for Pab1p (Dunn et al., 2005; Hilleren et al., 2001; Libri et al., 2002). However, the mechanism for the exosome mediated detection and retention of the defective transcripts at the transcription site is still far from being completely understood.

5.3 A final checkpoint at the nuclear periphery

Recent work has uncovered a quality control checkpoint located at the nuclear face of the NPC that assures that only mature, fully processed mRNA molecules are exported from the nucleus (reviewed by Fasken and Corbett, 2005; Sommer and Nehrbass, 2005). The key players in this checkpoint are proteins localised in the inner basket of the NPC known as myosin-like proteins (Mlp1 and Mlp2) in *S. cerevisae* and as translocated promoter region (Tpr) protein in vertebrates. The first evidence that these proteins might be important for mRNP export came from the observation that overexpression of either Mlp1p or Tpr results in the accumulation of mRNA in the nucleus. Further studies showed that Mlp1p interact with the mRNP component Nab2p and suggested that it might act in a checkpoint at the nuclear pore by interacting with export-competent mRNPs (Green et al., 2003). In agreement with this suggestion it was demonstrated that Mlp1p is involved in the retention of unspliced mRNAs in the nucleus (Galy et al., 2004). Deletion of *MLP1* allows unspliced transcripts to leak to the cytoplasm but has no effect on splicing. It was previously shown that proper retention on intron-containing mRNA molecules in the nucleus requires an intact 5’ splice site and

branchpoint (Legrain and Rosbash, 1989; Rain and Legrain, 1997). In agreement with this, an intact 5' splice site was also shown to be required in the case of Mlp1-based retention of pre-mRNA. The simplest explanation for this is that a factor that is dependent on this sequence for RNA binding is recognised by Mlp1p and serves as a marker of immature transcripts that should not be exported (Galy et al., 2004). Interestingly, the same authors identified an RNA-dependent interaction between Mlp1p and the branchpoint binding protein SF1, which suggests an interaction between Mlp1p and the intron containing transcripts (Galy et al., 2004). A recent paper adds further support for an important role of Mlp proteins in nuclear mRNA metabolism (Vinciguerra et al., 2005). This study shows that Mlp1p and Mlp2p retain intronless mRNP complexes produced in an Yra1 mutant background, that is, mRNPs with early assembly defects. More specifically, the loss of Mlp1p or Mlp2p increases the levels of transcripts that were reduced in Yra1 mutant strains and substantially rescues the growth defect of both Yra1 and Nab2 mutants but not of other mRNA export mutants. These results are consistent with a role of Yra1p and Nab2p in proper mRNP docking to Mlp proteins. This study also showed that mRNP complexes loaded with mutant Yra1p are sequestered on the Mlp platform and unable to proceed along the export pathway. This block negatively impacts on the expression of a subset of genes potentially located at the nuclear periphery (Vinciguerra et al., 2005).

6. The human β–globin as a model to study quality control of gene expression

Many of the most exciting discoveries in the field of eukaryotic gene expression are linked to globin research. The globin gene was the first gene to be cloned back in 1976, its promoter the first one to be extensively characterised in its constitutive elements and the so-called locus control region (LCR) represented the first demonstration of the existence of distant regulatory elements controlling complex clusters of genes (for a review see Cao and Moi, 2002). NMD was the first reported example of a quality control mechanism of gene expression, and the beneficial effects of this surveillance mechanism in the phenotype of inherited human diseases started to be appreciated in the case of β-thalassemia (reviewed by Khajavi et al., 2006).

The diseases caused by mutations in the globin genes, known as hemoglobinopathies, are the most common and diverse group of genetically inherited disorders. The disease phenotypes can be divided into two categories. The structural hemoglobinopathies, that arise from amino acid substitutions in the globin chains that result in the production of haemoglobin with structural defects. The best known disorder of this category is sickle cell anemia. The second category is the thalassemias, which result from a reduction in the production of α-globin (α-thalassemia) or β-globin (β-thalassemia) proteins with a consequent reduction or total absence of haemoglobin (see Olivieri, 1999). The absence of defective globin chains in this group of diseases and therefore their recessive phenotype is a consequence of quality control mechanisms of gene expression that operate at the RNA level to avoid the translation of defective protein products (reviewed by Custódio and Carmo-Fonseca, 2001).

6.1 The human β-globin gene cluster and the LCR

The human β-globin gene (*HBB*) cluster consists of five genes arranged in the short arm of chromosome 11 (region 11p15.5) in the same order in which they are expressed during development: 5'-ε-, Gγ -, Aγ-, δ-, and β-globin gene (Figure 8). The ε-globin gene is expressed during the first 6 weeks of embryonic life within the blood islands of the yolk sac; the two γ-globin genes are then expressed in the fetal liver throughout mid-pregnancy and start to decline just before birth. Finally, the adult β-globin gene begins to be expressed in the bone marrow around the time of birth (reviewed by Levings and Bungert, 2002; Li et al., 2002). Transcription in the *HBB* cluster is under the control of a distal regulatory element known as locus control region (LCR) which is located up to 50 kb upstream of the β-globin gene. The LCR contain four erythroid specific DNase I hypersensitive sites (HS) and a further upstream site (HS-5) (Figure 8). HS are a hallmark of DNA-protein interactions and of an "open" chromatin structure that facilitates the access of regulators and lower the threshold for activation of the linked genes. The LCR was first identified by its ability to confer position-independent, high level expression on a linked gene in the erythroid lineage of transgenic mice (Grosveld et al., 1987) and in erythroid cells in culture (Blom van Assendelft et al., 1989; Talbot et al., 1989). The level of expression of a transgene in transgenic mouse lines or cloned cells is usually low and variable. It is believed that the levels of expression of transgenes are influenced by the site of integration and the surrounding chromatin and the

LCR abolishes this "position effects" allowing expression levels directly proportional to transgene copy number (Grosveld et al., 1987). An alternative interpretation for the observed low levels of expression of transgenes is that the insertion of multiple-copy arrays at single sites may result in condensation into heterochromatin and transcriptional inactivation (Garrick et al., 1998). In this context, the LCR may function to abolish heterochromatinisation of the array, therefore allowing expression levels directly proportional to transgene copy number. This hypothesis is supported by the study of LCR function at its endogenous location in cell lines suggesting that its primary role in this context is not the establishment and maintenance of an open β-globin chromatin domain, but rather to function in globin gene transcription activation (Epner et al., 1998; Reik et al., 1998).

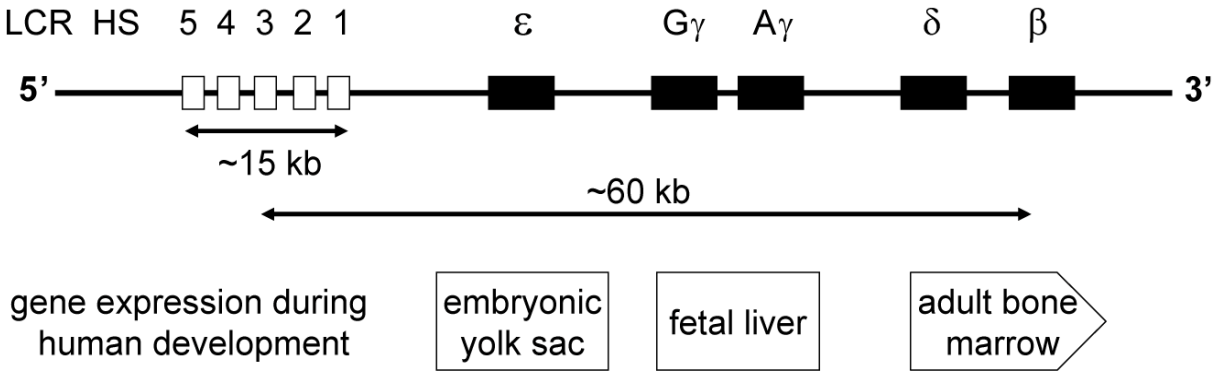


Figure 8 - The human β-globin gene locus (*not drawn to scale*). This locus consists of 5 functional genes, indicated as dark boxes, arrayed in their order of developmental expression, 5'-ε-Gγ-Aγ-δ-β-3'. There are 2 developmental switches in globin chain synthesis coincident with changes in site and type of erythropoiesis. During primitive erythropoiesis, the ε-globin gene is expressed in the embryonic yolk sac. The first switch occurs at approximately 8 weeks of gestation; the ε-globin gene is silenced and the Gγ- and Aγ-globin genes are expressed during definitive erythropoiesis in the fetal liver. The second switch occurs shortly after birth; the γ-globin genes are silenced and the β-globin gene and, to a lesser extent, the δ-globin gene are activated in the bone marrow. The locus control region (LCR) is represented as the sum of the five DNase I hypersensitive (HS) sites, indicated as white boxes. It should be noted that additional HS sites were mapped 5' to HS-5, but it is currently not known whether these sites participate in globin gene regulation or whether they are associated with the regulation of genes located upstream of the globin locus. The HS core elements are 200-400 bp in size and separated from each other by more than 2 kb. Adapted from (Levings and Bungert, 2002).

Several models have been proposed over the years to explain the LCR mode of action. Among them, the looping and tracking/scanning models are the most prevalent (Figure 9) (reviewed by Li et al., 2002). The looping model (Figure 9A) proposes that the LCR acts as an integral unit (a "holocomplex") to interact directly with individual globin genes by looping

through the nucleoplasm and stimulating their transcription (Bungert et al., 1995; Choi and Engel, 1988; Wijgerde et al., 1995). This model can explain the balanced competition between the genes and the LCR for function (reviewed by Grosveld et al., 1998), the fact that the LCR can activate only one gene at a time and the flip-flop transcription observed in the β-globin locus (Wijgerde et al., 1995). This flip-flop was observed using *in situ* hybridisation that allowed the detection of primary γ-globin transcripts in the nucleus of individual embryonic cells that had already switched and display predominantly β-globin chain in the cytoplasm. This experimental evidence indicates that the LCR can flip-flop back and forth to activate adjacent genes such as the γ- and β-globin genes and that it is the total time spent by the LCR in each gene interaction that determines the prevailing chain in the cytoplasm (Wijgerde et al., 1995). Further support for this model came from the observations that HS of the LCR are in close physical proximity to an actively transcribed β-globin gene located over 50 kb away *in vivo* (Carter et al., 2002; Tolhuis et al., 2002). However, this model has not provided a mechanism to explain some aberrant transcripts detected across the LCR and intergenic regions in erythroid cells (Ashe et al., 1997). In the tracking or scanning model (Figure 9B), erythroid-specific and ubiquitous transcription factors and cofactors bind recognition sequences in the LCR, forming an activation complex that migrates, or tracks, linearly along the DNA helix of the locus (Blackwood and Kadonaga, 1998; Tuan et al., 1992). When this transcription complex encounters the basal transcription machinery located at the correct promoter (according to the developmental stage), the complete transcriptional apparatus is assembled and transcription of that gene is initiated. This model can explain the intergenic transcription that could function to deliver transcription complex proteins to the globin gene promoters via the tracking mechanism. A facilitated-tracking model that incorporates aspects of both the looping and tracking models was also proposed (Figure 9C) (Blackwood and Kadonaga, 1998). According to this hybrid model the LCR bound to transcription factor and coactivators loops to contact downstream DNA in promoter-distal regions, where the transcription factor complex is released. This complex then tracks in small steps along the chromatin until it encounters the appropriate promoter with its associated bound proteins, a stable loop structure is established and gene expression proceeds.

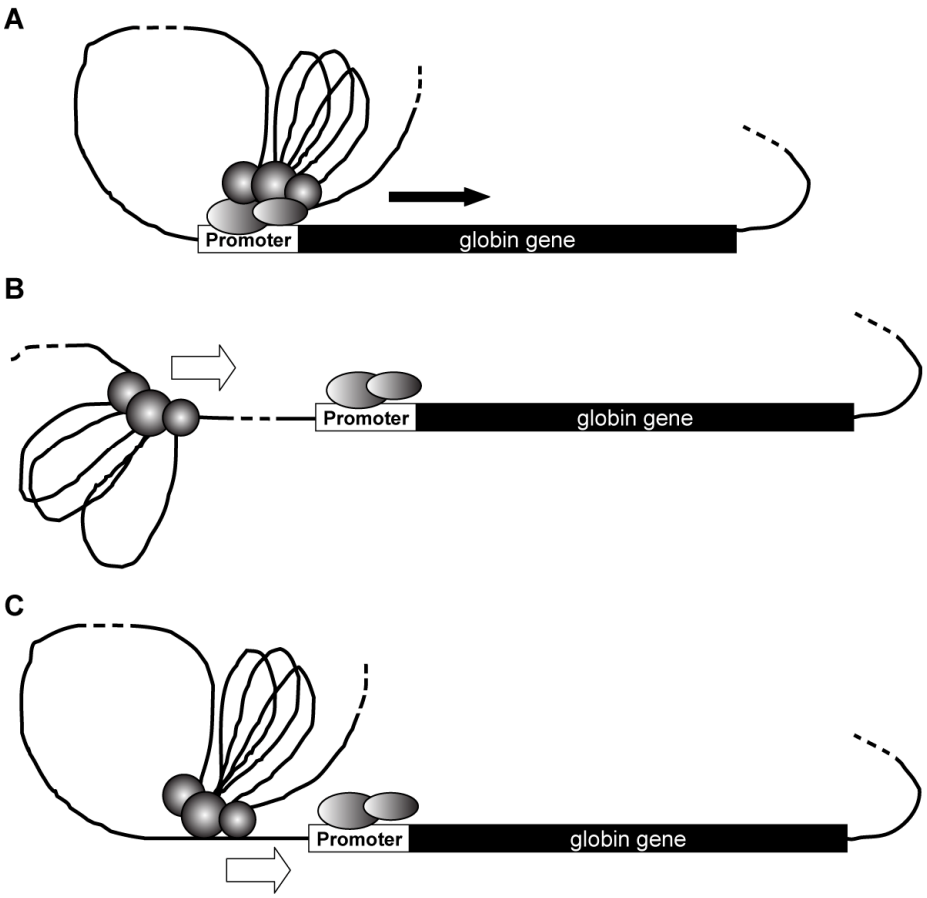


Figure 9 - Models of LCR function. Transcription factors are shown as ovals and circles. The flanking DNA sequences of the HSs are depicted as loops between the HS cores. **(A)** Looping model. Transcription factors bind to the LCR HSs and the gene promoter. The LCR directly interacts with the gene promoter by looping out the intervening DNA, thus forming an active transcription complex at the gene promoter. **(B)** Tracking or scanning model. Sequence-specific transcription factors bind to the LCR, forming a complex that tracks down the DNA sequence, as depicted by the large white arrow, until encountering transcription factors bound to the appropriate gene promoter, initiating high-level gene expression. **(C)** Facilitated tracking model. Aspects of both looping and tracking models are combined. Sequence-specific transcription factors bind the LCR, looping then occurs to deliver the bound transcription factors proximal to the gene promoter, followed by tracking, until they encounter transcription factors bound to the appropriate gene promoter. Adapted from (Li et al., 2002).

6.2 The LCR/MEL expression system

Since the discovery of the LCR, a great number of HS dissections and combinations have been done in different experimental systems to identify the minimal sequence requirements to create an LCR compact enough to fit available vectors for the gene therapy of the thalassemias and sickle cell disease. During the course of this line of research a shorter

version of the LCR (with only 6.5 kb instead of the normal 20 kb) has been engineered containing sequences from the HS sites 1-4. This micro-locus control region (βLCR) was shown to replace the role of the LCR in transgene expression (Collis et al., 1990). Eukaryotic expression vectors where developed taking advantage of this βLCR, in which the gene of interest is placed under the control of the *HBB* promoter and cloned downstream of the βLCR. Stable transfection of these vectors in erythroid cells (because the globin LCR is erythroid specific) has allowed high levels of expression of transgenes which are directly proportional to the transgene copy number (Needham et al., 1992). This makes possible quantitative comparisons for example between wild-type genes and those carrying different mutations.

The type of erythroid cells that has been extensively used as an expression system with this kind of vectors are the murine erythroleukaemia (MEL) cells. MEL cells are murine virus-transformed spleen hematopoietic cells, which grow in suspension culture. These progenitor cells, arrested at the pro-erythroblast stage of differentiation, can be induced in culture with various chemical agents to undergo terminal erythroid differentiation. MEL cells committed to the erythroid differentiation programme go through a series of morphological and biochemical alterations that resemble those seen in the differentiation of normal early erythroid progenitors. These include transcriptional activation of several genes related to the erythroid phenotype and selective repression of many others, progressive changes in cell surface architecture, condensation of nucleus, reduction of cell size and even nuclear extrusion under certain culture conditions (for an extensive review see Tsiftsoglou et al., 2003). Among the genes exclusively related to the erythroid phenotype, whose expression is induced upon differentiation, are the murine globin genes (α-globin, β^{major}-globin and β^{minor}-globin), resulting in very high levels of haemoglobin in the cytoplasm that give a pink-red colour to the differentiated cell pellets. In MEL cells transfected with βLCR expression vectors the expression of the transgenes under the control of the βLCR parallel the expression of the endogenous murine globin genes.

6.3 Using the LCR/MEL expression system to study quality control of gene expression in the nucleus

HBB genes defective in some aspects of pre-mRNA processing have been expressed under the control of βLCR in MEL cells (Antoniou, 1991). This type of research initially demonstrated that the larger, second intron of *HBB* is absolutely required for the accumulation

of cytoplasmic mRNA (Collis et al., 1990). Transgenes lacking this intron give rise to nuclear localised transcripts which are not correctly 3' end cleaved (Antoniou et al., 1998). The expression of an *HBB* gene with a mutated 5' splice site at the second intron results in very low levels of β-globin mRNA in the cytoplasm, a phenotype identical to that seen in β-thalassemia patients harbouring similar mutations (Antoniou et al., 1998). Given this initial observation that *HBB* transcripts defective in either splicing or 3' end cleavage were not detected in high levels in the cytoplasm of MEL cells, we hypothesised that some kind of quality control mechanism linked to the processing of the primary transcripts should be operating in the nucleus. The βLCR expression system in MEL cells offered us great potential to explore this hypothesis in an *in vivo* context. The high level of gene expression afforded by the βLCR in this system made it easier to develop an *in situ* hybridisation approach to study the nuclear localisation of both wild-type and processing defective mutant transcripts to start to explore the quality control mechanisms of gene expression in the mammalian cell nucleus.

7. Goals of the work

The main goal of this work was to elucidate the quality control mechanism responsible for the retention within the nucleus of human β-globin transcripts with mutations that impaired their processing. To reach this goal we proposed to:

1- Determine the intranuclear localisation of the retained transcripts.

2- Identify the molecular players responsible for the retention.

CHAPTER II

RESULTS

1. Inefficient processing impairs release of RNA from the site of transcription

Initial biochemical studies have shown that human β-globin (*HBB*) transcripts with mutations that impaired their processing were underrepresented in the cytoplasm of MEL cells and enriched in the nuclear fraction, a phenotype identical to that seen in β-thalassemia patients harbouring similar mutations (Antoniou et al., 1998). This result was consistent with the retention in the nucleus of incorrectly processed transcripts. Based on this initial observation, we hypothesised the existence in the mammalian nucleus of a quality control mechanism to retain the incorrectly processed transcripts. To elucidate the mechanism responsible for the observed retention the first goal was to determine the intranuclear localisation of the retained transcripts. To address this, our experimental approach was based on the direct visualisation of both normal and mutated *HBB* transcripts using fluorescence *in situ* hybridization and confocal microscopy. The principal conclusion from this part of the work was that the intranuclear retention of the abnormally processed transcripts was taking place in close proximity to the site of transcription.

Work published in: The EMBO Journal Vol.18 No.10 pp.2855-2866, 1999

I would like to stress that some of the results presented and discussed in this section are the product of collaborative work. Dr. Michael Antoniou, Dr. Finola Geraghty and Dr. Frank Grosveld were responsible for the generation of MEL cell clones analysed in this work. The statistical analysis of the results was performed by Dr. H. Sofia Pereira.

The EMBO Journal Vol.18 No.10 pp.2855–2866, 1999

Inefficient processing impairs release of RNA from the site of transcription

Noelia Custodio, Maria Carmo-Fonseca, Finola Geraghty[1], H.Sofia Pereira, Frank Grosveld[2] and Michael Antoniou[1,3]

Institute of Histology and Embryology, Faculty of Medicine, University of Lisbon, 1699 Lisbon codex, Portugal, [1]GKT Medical School, Department of Experimental Pathology, Guy's Hospital, London SE1 9RT, UK and [2]MGC Department of Cell Biology and Genetics, Erasmus University, 3015 GE Rotterdam, The Netherlands

[3]Corresponding author
e-mail: michael.antoniou@kcl.ac.uk

We describe here for the first time the site of retention within the nucleus of pre-mRNA processing mutants unable to be exported to the cytoplasm. Fluorescence *in situ* hybridization was used to detect transcripts from human β-globin genes that are either normal or defective in splicing or 3′ end formation. Nuclear transcripts of both wild-type and mutant RNAs are detected only as intranuclear foci that colocalize with the template gene locus. The kinetics of transcript release from the site of transcription was assessed by treatment of cells with the transcriptional inhibitors actinomycin D, α-amanitin and DRB. These drugs induce the rapid disappearance of nuclear foci corresponding to wild-type human β-globin RNA. In contrast, pre-mRNA mutants defective in either splicing or 3′ end formation and which fail to be transported to the cytoplasm, are retained at the site of transcription. Therefore, 3′ end processing and splicing appear to be rate limiting for release of mRNA from the site of transcription.

Keywords: pre-mRNA processing/RNA transport/transcription

Introduction

Transport of mRNA from the nucleus to the cytoplasm is essential for the expression of eukaryotic genes. It is an active and highly selective process that involves *cis*-acting signals and specific *trans*-acting factors (for recent reviews see Lee and Silver, 1997; Nakielny *et al.*, 1997; Nigg, 1997). Although much recent information has begun to define the pathways that mediate and control the nucleocytoplasmic RNA traffic in a cell, very little is known about the release of mRNAs from the site of transcription for subsequent transport to the nuclear periphery and translocation across the nuclear pore.

Several lines of evidence indicate that efficient mRNA transport involves cotranscriptional interaction with RNA-binding proteins and a correct processing of the pre-mRNA into mature mRNA (Izaurralde and Mattaj, 1992; Elliot *et al.*, 1994; Visa *et al.*, 1996a,b). In particular, a relationship between splicing and export has been clearly established (Chang and Sharp, 1989; Legrain and Rosbash, 1989; Hamm and Mattaj, 1990). The removal of introns from pre-mRNA is catalysed by the spliceosome, a dynamic complex composed of small nuclear ribonucleoproteins (snRNPs) and numerous protein components (Moore *et al.*, 1993; Kramer, 1996). Spliceosome assembly occurs on each substrate pre-mRNA *de novo* and requires conserved recognition sequences located at the exon–intron boundaries. Pre-mRNA molecules bearing mutations that allow spliceosome assembly but impair splicing are largely retained in the nucleus, whereas mutations that disturb splicing complex formation can partially overcome the block in transport of intron-containing pre-mRNAs to the cytoplasm (Chang and Sharp, 1989; Legrain and Rosbash, 1989; Hamm and Mattaj, 1990). This led to the hypothesis that spliceosome assembly may cause retention of pre-mRNA in the nucleus, either because certain splicing factors interact with nuclear structures holding the unspliced RNA, or because the spliceosome may prevent interaction of the RNA with the export machinery (Nakielny *et al.*, 1997). In addition to splicing, 5′ capping and 3′ end formation are also known to influence mRNA export. In general, both the 5′ cap and the 3′ poly(A) tail enhance the export rate of a transcript but they do not appear to be essential (Eckner *et al.*, 1991; Jarmolowski *et al.*, 1994). However, mutated β-globin RNAs devoid of the second intron are unable to undergo correct 3′ end formation and are retained in the nucleus (Collis *et al.*, 1990; Antoniou *et al.*, 1998).

Spliceosome assembly and splicing are known to occur during ongoing transcription elongation *in vivo* (Beyer and Osheim, 1988; LeMaire and Thummel, 1990; Bauren and Wieslander, 1994; Tennyson *et al.*, 1995). Accordingly, spliceosome components and spliced RNAs have been visualized in close proximity to sites of transcription (Wu *et al.*, 1991; Xing *et al.*, 1993; Zhang *et al.*, 1994; Bauren *et al.*, 1996; Huang and Spector, 1996; Neugabauer and Roth, 1997). In addition, it has recently been discovered that the splicing and polyadenylation machinery can associate with the transcription elongation complex via the carboxy-terminal domain (CTD) of the large subunit of RNA polymerase II (Steinmetz, 1997). However, it is not clear from these studies whether pre-mRNA processing must precede release for subsequent transport to the nuclear pores. Also, there is currently no information available on the kinetics of processing and release of mRNA from the site of transcription.

In order to address these questions we have used *in situ* hybridization and confocal microscopy to visualize the release of normal and mutated human β-globin RNAs from the vicinity of their cognate gene templates within the nucleus. The system we have exploited in these studies consists of human β-globin genes under the control of the locus control region (LCR) that are stably transfected

into murine erythroleukaemia (MEL) cells. The human β-globin genes within this context reproducibly express at physiological levels which are directly proportional to transgene copy number and independent of the site of integration in the host cell genome during the induced terminal differentiation of these cells (Blom van Assendelft *et al.*, 1989; Talbot *et al.*, 1989; Collis *et al.*, 1990). This pattern of gene expression demonstrates that although the transgenes are integrated at ectopic sites within the host cell genome as a tandem array, their function accurately reflects that observed from the same genes at their native chromosomal locus. Even globin loci integrated in heterochromatin regions are transcribed and processed at normal levels (Milot *et al.*, 1996).

Our data show that inhibition of transcription by actinomycin D, α-amanitin or 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB) causes the rapid release of wild-type human β-globin RNA from the vicinity of the site of transcription. In contrast, mutant globin pre-mRNAs that are defective in either splicing or 3′ end formation are held in close proximity to the gene template in the presence of these drugs. These observations imply that efficient pre-mRNA processing is crucial and therefore rate limiting for the release of transcripts from the site of transcription.

Results

Localization of human β-globin RNA in MEL cells

The wild-type human β-globin gene (βWT) within the microlocus LCR expression cassette (Figure 1A; Collis *et al.*, 1990) was transfected into MEL cells and stable clones or populations selected with G418, as described (Antoniou, 1991). One of these clones, MELβWT, which harbours ~14 copies of the transgene as a tandem array (data not shown), was used to perform *in situ* hybridization experiments. Expression of the human β-globin transgenes was induced by adding dimethylsulfoxide (DMSO) to the culture medium for 1–4 days during which the MEL cells undergo terminal erythroid differentiation (Antoniou, 1991; Antoniou *et al.*, 1993). Uninduced and induced cells were hybridized under non-denaturing conditions with a probe complementary to the transcribed sequence of the human β-globin gene (RNA probe, Figure 1A). Uninduced cells were devoid of any hybridization signal (data not shown), whereas cells induced for 4 days contain a fluorescent focus in the nucleus and additional cytoplasmic staining (Figure 1B, a–c). The intensity of staining of the nuclear foci reached maximum levels after 2 days of induction whereas the cytoplasmic labelling peaked after 4 days of differentiation (data not shown). This is consistent with biochemical data showing that the rate of globin gene transcription in MEL cells attains maximum levels 36–48 h after induced differentiation and cytoplasmic accumulation peaks after 4 days (see Antoniou, 1991). Labelling of total cellular DNA with TO-PRO-3 confirmed the intranuclear localization of the human β-globin RNA foci (Figure 1B, b). In addition, superimposition of confocal optical sections revealed that intranuclear β-globin RNA is detectable only as a single focus within the nucleus with no other sites of accumulation (Figure 1B, c).

In order to determine whether the RNA foci observed in the nuclei of induced cells represent sites of transcription of the transfected human β-globin genes, double-hybridization experiments were performed. Cells were sequentially hybridized for human β-globin RNA under non-denaturation conditions with a probe complementary to the transcribed sequence of the gene (RNA probe; Figure 1A), followed by detection of the corresponding gene locus by hybridizing under denaturing conditions with a probe complementary to the plasmid cassette used for transfection (DNA probe; Figure 1A). As expected for a clonal population of stably transfected cells, hybridization with the DNA probe produces a fluorescent focus in each nucleus (Figure 1B, e). In contrast, RNA foci are only detected in ~20–40% of the cells (Figure 1B, d). A similar result is observed when immunofluorescence is performed using an antibody specific for human β-globin (data not shown; see Fraser *et al.*, 1993), indicating that expression of human β-globin is restricted to a subset of the transfected population. This is probably due to a combination of the asynchronous nature of the cell cultures and position–effect variegation (see Milot *et al.*, 1996). Nevertheless, the overlay of red (DNA hybridization) and green (RNA hybridization) images shows that the focal signals overlap in the nucleus (Figure 1B, f). Given the limits of resolution of light microscopy, these results indicate that the foci labelled by the RNA probe are likely to correspond to the sites of human β-globin gene transcription.

Since the RNA probe is complementary to the full-length β-globin transcript, it does not allow nascent and terminating or terminated transcripts to be distinguished. Therefore, to identify specifically those RNA molecules that have been elongated towards the end of the transcription unit and that are either terminating or terminated, we used a 3′ RNA probe (Figure 1A; see Materials and methods). The 3′ RNA probe spans the poly(A)-addition site and will, as a result, only hybridize to those β-globin RNAs that have been extended to this terminal region of the gene. In addition to cytoplasmic labelling, this probe produces a focal intranuclear signal which colocalizes with the focus produced by the full-length RNA probe (Figure 1B, g–i). These data show that our *in situ* hybridization procedure has the sensitivity to detect terminating or terminated β-globin RNA chains at the site of transcription.

β-globin pre-mRNA is spliced at the site of transcription

We next asked whether splicing of β-globin pre-mRNA occurs while the transcripts are still in the vicinity of the site of transcription. In order to address this question, splice junction oligonucleotide probes (SJ I/II and SJ II/III), which are complementary to the ends of the three exons and span the two introns of the human β-globin gene (Figure 2A), were employed. The oligonucleotides used were the same as described previously (Zhang *et al.*, 1996) and had been shown to be incapable of forming stable hybrids with unspliced RNA. MELβWT cells that harbour the wild-type human β-globin gene were induced to differentiate and hybridized with a mixture of the SJ I/II and SJ II/III probes. The results (Figure 2B, a and d) show a cytoplasmic signal with additional labelling of intranuclear foci. Double-hybridization experiments using these splice junction oligonucleotide probes and the full-length RNA probe (Figure 1A) demonstrate that the signals

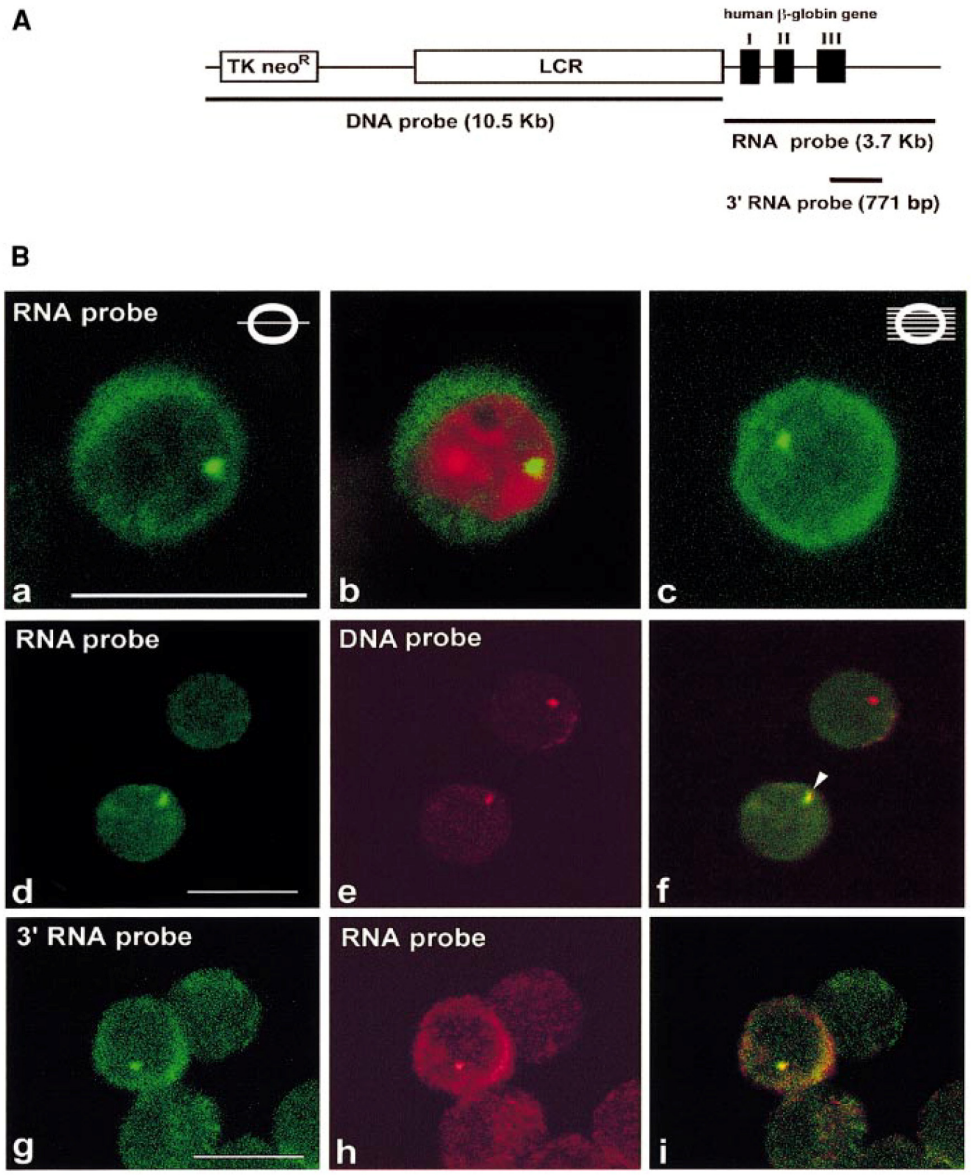
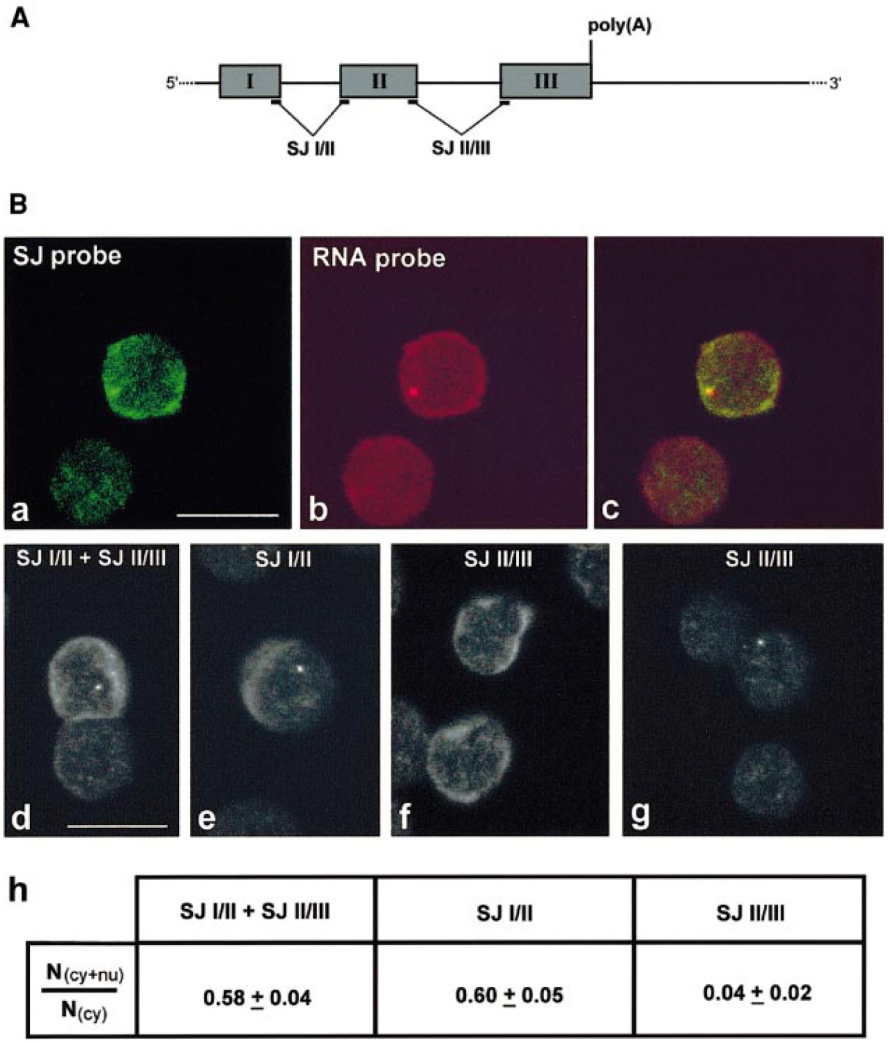


Fig. 1. (**A**) Schematic illustration of the human wild-type β-globin construct. The wild-type human β-globin gene (βWT) is within the micro-locus control region (LCR) expression cassette (Collis *et al.*, 1990; see Materials and methods). Human β-globin gene exons are shown as black rectangles. The TK *neo*^{R} gene confers resistance to G418 in stably transfected MEL cells. The extent of the probes used for *in situ* hybridization to detect human β-globin RNA (RNA probe; 3′ RNA probe) and site of transgene integration (DNA probe) are also shown. (**B**) *In situ* detection of human β-globin RNA in transfected MEL cells. MEL cells transfected with the βWT construct (MELβWT) were fixed either in formaldehyde and permeabilized with Triton X-100/saponin (**a**–**c**) or in formaldehyde/acetic acid and digested with pepsin (**d**–**i**). (a) and (b) depict a single confocal section through a cell hybridized with the RNA probe (2 ng/μl) that is complementary to the entire length of the human β-globin transcription unit (green staining); total DNA was labelled with the dye TO-PRO-3 (Molecular Probes) (red staining). (c) depicts a superimposition of 10 optical sections through another cell hybridized with the RNA probe (consecutive sections are separated by 0.7 μm). The cells were induced to undergo erythroid differentiation for 4 days. (d)–(f) depict the simultaneous detection of human β-globin RNA and the transfected gene locus. MELβWT cells were induced for 2 days, and hybridized with the RNA probe labelled with digoxigenin (4 ng/μl) (e, green staining). The cells were then fixed in formaldehyde, denatured and hybridized with the DNA probe labelled with dinitrophenyl (2 ng/μl) (d, red staining). Fluorescein- and Texas Red-coupled antibodies revealed the sites of hybridization of the RNA and DNA probes, respectively. Superimposition of red and green images shows that the DNA and RNA foci in the nucleus colocalize (f). (**g**)–(**i**) show sequential hybridization with the 3′ RNA probe (3 ng/μl) labelled with digoxigenin (revealed with fluorescein, g), and the full-length RNA probe (4 ng/μl) labelled with DNP (detected with Texas Red, h). (i) depicts a superimposition of the two images. Cells were induced for 4 days. Bar, 10 μm.



h

	SJ I/II + SJ II/III	SJ I/II	SJ II/III
$\frac{N_{(cy+nu)}}{N_{(cy)}}$	0.58 ± 0.04	0.60 ± 0.05	0.04 ± 0.02

Fig. 2. Spliced human β-globin RNAs are detected at the site of transcription. (**A**) Schematic illustration of the splice junction (SJ) probes. The SJ I/II probe is complementary to the last 12 nucleotides of exon I and the first 12 nucleotides of exon II, whereas the SJ II/III probe hybridizes to the last and first 12 nucleotides of exons II and III, respectively. (**B**) MELβWT cells were induced for 2 days, fixed in formaldehyde/acetic acid, digested with pepsin and hybridized with a mixture of splice junction oligonucleotide probes (SJ I/II and SJ II/III, 1 ng/μl each) labelled with DNP and revealed with fluorescein (**a**). The cells were then fixed in formaldehyde and hybridized with the full-length RNA probe (4 ng/μl) labelled with digoxigenin and revealed with rhodamine (**b**). (**c**) depicts a superimposition of the two images. In (**d**)–(**g**) MELβWT cells were induced for 4 days, fixed in formaldehyde/acetic acid, digested with pepsin and hybridized with either a mixture of the splice junction oligonucleotide probes SJ I/II and SJ II/III (d), SJ I/II (e) or SJ II/III (f). As a positive control, MELβIVSI cells were induced for 4 days and hybridized with the SJ II/III probe (g). The proportion of labelled cells containing nuclear foci was estimated according to the formula: [*N*(cy + nu)/*N*(cy)]. (**h**) depicts means ± SE (three separate experiments were performed for each probe and a total of 100–200 cells were analysed). Bar, 10 μm.

colocalize in the nucleus (Figure 2B, b and c), indicating that splicing of globin pre-mRNA is taking place in close proximity to the site of transcription. However, different results were obtained when the splice junction oligonucleotide probes were used separately. The SJ I/II probe which spans exons I and II, produces an intranuclear hybridization signal (Figure 2B, e) similar to that obtained with the mixture of the two probes (Figure 2B, d). This indicates that splicing of intron I is taking place in the vicinity of the site of transcription. In contrast, the exon II–III spanning SJ II/III probe fails to label nuclear foci in the vast majority of cells (Figure 2B, f). In the cytoplasm, both probes produce a strong hybridization signal (Figure 2B, d–f). Potential technical artefacts with the SJ II/III probe were controlled for by hybridization to MELβIVSI cells. These cells harbour a mutant human β-globin gene which lacks the second intron and therefore possesses a fusion of exons II and III (Figure 4B, upper panel; see below). MELβIVSI cells hybridized with SJ II/III show clearly labelled intranuclear foci (Figure 2B, g) demonstrating that this probe is functioning normally. (Note the absence of cytoplasmic labelling in MELβIVSI cells,

consistent with the finding that this mutant RNA fails to be exported from the nucleus; Antoniou *et al.*, 1998 and see below.)

A quantitative analysis of wild-type MELβWT cells hybridized with the splice junction SJ probes either separately or in combination was conducted (Figure 2B, h). The proportion of cells that are labelled in the cytoplasm and also contain a nuclear focus is similar for the mixture of the two probes (58%) or the exon I–II spanning SJ I/II probe alone (60%). In contrast, the majority (96%) of cells labelled in the cytoplasm by the exon II–III spanning SJ II/III probe is devoid of a detectable intranuclear focus. These data suggest that the first intron of human β-globin RNA is spliced while the transcript is still at the gene locus, whereas removal of the second intron takes place either immediately prior to rapid release from the site of transcription or at some other location within the nucleus after transport from the region of the gene template.

Actinomycin D causes a rapid release of β-globin transcripts from the site of transcription

The observed intranuclear foci of β-globin RNA appear to represent newly synthesized transcripts in the vicinity of the gene locus. If this is indeed the case, treating cells with transcription inhibitors should lead to the disappearance of these nuclear foci as a result of transport away from the site of transcription of previously synthesized RNA molecules. This in turn may provide insight into the kinetics of transcript release from the site of transcription. In order to test this idea, we initially used actinomycin D, a drug that acts very rapidly *in vivo* (Darnell *et al.*, 1971) and exerts its effects by binding to the DNA template, thereby interfering with the elongation of the growing RNA chain (Kersten *et al.*, 1960; Goldberg *et al.*, 1962).

The results show that when MELβWT cells are treated with actinomycin D for 5 min and hybridized with the full-length RNA probe (Figure 1A), the intranuclear foci are no longer detected (Figure 3a and b). However, cytoplasmic labelling remains visible, confirming that these cells were transcribing the transfected, wild-type human β-globin genes before exposure to the drug (Figure 3b). Quantitative analysis reveals that actinomycin D causes a highly significant decrease in the proportion of cells containing a visible focus in the nucleus (Figure 5A, MELβWT). Similar results were obtained with the 3′ RNA probe (Figure 1A) during a time course experiment which shows that within 1 min of exposure to the drug, the proportion of cells with a visible focal signal in the nucleus decreases to approximately one-third (Figure 3c). Thus, in the presence of actinomycin D, β-globin RNAs have a half-life of <1 min at the site of transcription.

β-globin RNA processing mutants are retained at the site of transcription

The data presented thus far establish that actinomycin D induces a rapid release of newly synthesized β-globin RNA from the site of transcription (Figure 3). We next analysed the effect of this drug on RNA processing mutants defective in cytoplasmic transport.

We have previously generated stably transfected MEL cell pools harbouring a human β-globin gene possessing a 5′ splice site mutation (GT→AC) of the second intron

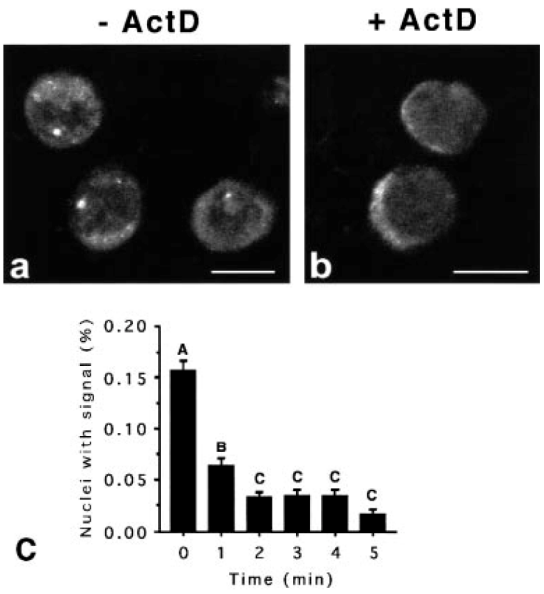


Fig. 3. Actinomycin D induces a rapid release of β-globin RNA from the site of transcription. (a) and (b) MELβWT cells were induced to undergo erythroid differentiation for 2 days and treated with actinomycin D. The cells were fixed in formaldehyde, permeabilized with Triton X-100/saponin and hybridized with the full-length or 3′ RNA probes. (**a**) Cells untreated with actinomycin D. (**b**) Cells treated with actinomycin D for 5 min before fixation and hybridization. Bar, 10 μm. (**c**) Kinetics of release of β-globin RNA from the site of transcription in the presence of actinomycin D. MELβWT cells were induced for 2 days, fixed in formaldehyde/acetic acid, digested with pepsin and hybridized using the 3′ RNA probe. The cells were either untreated (time 0) or treated with actinomycin D for 1, 2, 3, 4 or 5 min. The proportion of cells with a nuclear RNA signal was estimated in three independent experiments for each time point. A total of 12 microscopic fields corresponding to a total of 300–400 cells were analysed per experiment. There were no significant differences between experiments, allowing them to be pooled. The mean proportion of cells with nuclear foci is plotted (mean ± SE) and the values compared using a one-way analysis of variance, ANOVA (SAS Institute, 1990). A significant decrease in the mean proportion of positive cells over the 5 min period is observed [$F_{(5239)} = 88.72$, $p < 0.0001$]. In order to determine which time slots differed significantly, a Student–Newman–Keuls (SNK; SAS Institute, 1990) was used as an *a posteriori* test. This showed three groupings of values (SNK, $p < 0.05$). At time 0 the mean proportion of positive cells is significantly greater than at all other time slots; the mean value at time 1 is also significantly greater than the remaining time slots and there are no significant differences between mean values at times 2, 3, 4 and 5. Different letters above the histogram bars are used to represent statistically significant differences between means.

(Figure 4A, upper panel; Antoniou *et al.*, 1998). Biochemical analysis indicates that *in vivo* this mutant RNA (here referred to as *'*single splice site mutant*'*, βSM), is correctly 3′ cleaved and polyadenylated at normal rates but is not spliced and not transported to the cytoplasm (Antoniou *et al.*, 1998). In addition, *in vitro* this same mutant RNA is able to support at least partial spliceosome assembly (Lamond *et al.*, 1987). In the present study, a clone transfected with βSM (MELβSM) was used for *in situ* hybridization. After 4 days of induced erythroid differentiation, hybridization of MELβSM cells with the full-length RNA probe (Figure 1A) reveals one focal signal per nucleus (Figure 4A, a). However, no cytoplasmic staining is detected, consistent with the biochemical data

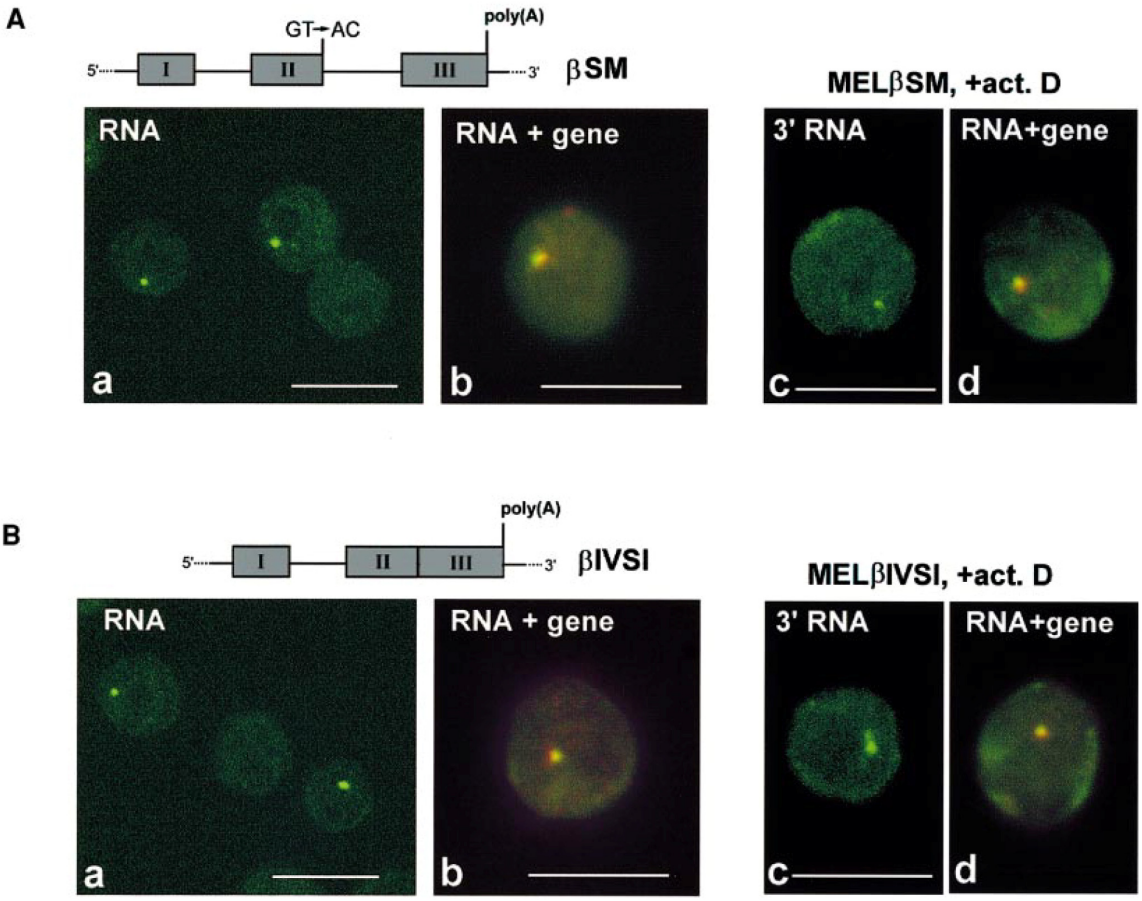


Fig. 4. β-globin RNA mutants for splicing and 3′ end processing are retained near the site of transcription. (**A**) MELβSM cells transfected with the construct βSM that contains a 5′ splice donor site mutation (GT→AC) of the second intron (upper illustration) were induced to undergo erythroid differentiation with DMSO for 4 days. (**a**) Hybridization with the RNA probe. Note that a focal signal is readily visible in the nucleus, but no cytoplasmic staining is detected. (**b**) Double-labelling of a cell with the RNA and DNA probes. Note that the RNA focus is localized in close vicinity to the site of transcription. (**c**) and (**d**) MELβSM cells treated with actinomycin D for 5 min. Hybridization with the 3′ RNA probe reveals an intranuclear focus (c), and double-hybridization using the RNA and DNA probes shows that the RNA focus colocalizes with the site of transcription (d). (**B**) A similar set of experiments was performed using MELβIVSI cells, which contain a β-globin construct that lacks completely the second intron (IVS-II), possessing only IVS-I (βIVSI; upper panel illustration). Cells were induced with DMSO for 4 days. (**a**) Hybridization with the RNA probe shows intranuclear foci but no cytoplasmic staining. (**b**) Double-hybridization using the RNA and DNA probes confirms that the RNA focus is localized near the site of transcription. (**c**) and (**d**) After treatment with actinomycin D for 5 min, hybridization with the 3′ RNA probe reveals the presence of intranuclear foci (c) and double-hybridization demonstrates that the RNA is retained near the site of transcription (d). Bar, 10 μm. Note: no cytoplasmic staining is seen in any of the panels depicted as the transcripts from these mutant human β-globin genes are defective in transport. The staining observed at the rim of the nuclei shown in (A, c/d and B, d) is due to non-specific background hybridization and trapping of the probe. This is shown by the fact that a similar pattern can be seen when the same probes are used with untransfected, uninduced, (negative control) MEL cells (data not shown).

demonstrating that these RNA molecules fail to be exported from the nucleus (Antoniou *et al.*, 1998). Double-hybridization experiments show that the foci corresponding to this mutant βSM RNA, colocalize with the gene template (Figure 4A, b).

In contrast with the results obtained with the wild-type βWT construct (Figure 3), focal signals of 5′ splice site mutant βSM RNA remain visible in the nuclei of cells treated with actinomycin D for 5 min (Figure 4A, c and d; Figure 5A, MELβSM). Furthermore, the βSM RNA foci colocalize with the signal produced by DNA hybridization for the transgene template, suggesting that these mutant transcripts are not released from the vicinity of the site of transcription (Figure 4A, d).

As 3′ end processing also plays an important role in transport of β-globin RNA to the cytoplasm (Collis *et al.*, 1990; Antoniou *et al.*, 1998), we extended our analysis to cells transfected with a construct that lacks completely the second intron (IVS-II), possessing only IVS-I (βIVSI; Figure 4B, upper panel). Despite possessing normal 3′ processing signal sequences, this mutant is unable to undergo correct 3′ end formation, producing an RNA species that is not cleaved and fails to reach the cytoplasm (Collis *et al.*, 1990; Antoniou *et al.*, 1998). The *in situ* hybridization results show that βIVSI RNA is detected in close proximity to the site of transcription but not in the cytoplasm (Figure 4B, a and b). In addition, intranuclear foci of βIVSI mutant transcripts that colocalize with

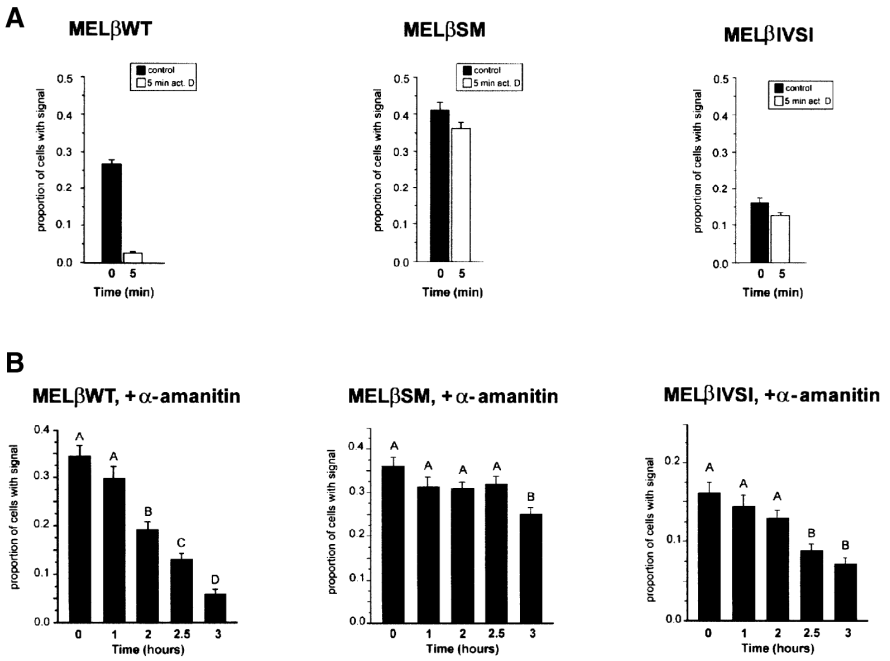


Fig. 5. Quantitative analysis of the effect of transcription inhibitors on β-globin RNA release. (**A**) MELβWT, MELβSM and MELβIVSI cells were induced to undergo erythroid differentiation for 2 days, untreated or treated with actinomycin D for 5 min and hybridized with the full-length RNA probe. The proportion of cells with nuclear foci was assessed within ~10 randomly selected microscopic fields corresponding to a total of 300–400 cells per experiment. Three separate experiments were conducted which showed no significant difference between sets of data allowing values to be pooled. Means ± SE are plotted. The values were compared using a Student's *t*-test (SAS Institute, 1990). Treatment with actinomycin D induces a highly significant decrease in the mean proportion of MELβWT cells with a hybridization signal in the nucleus [Student's *t*-test, $t_{(48.6)} = 18.94$, $p = 0.0001$] but not with MELβSM cells [$t_{(68)} = 2.02$, $p = 0.05$]. MELβIVSI cells show a slight decrease [$t_{(51.2)} = 2.43$, $p = 0.02$) in the mean proportion of nuclei with a signal in the presence of this drug. (**B**) The effect of α-amanitin on release of human β-globin transcripts from the site of transcription was studied on MELβWT, MELβSM and MELβIVSI cells that had been induced for 2 days. The statistical analysis was performed as described in the legend to Figure 3c. Two independent experiments were conducted for each cell type. In each experiment, 10 microscopic fields were counted for every time point. There were no significant differences between experiments, allowing them to be pooled. Statistically significant different means within each cell type (SNK, $p < 0.05$) are represented by different letters.

the gene locus remain visible following treatment with actinomycin D for 5 min (Figures 4B, c and d, and 5A, MELβIVSI).

Therefore, in marked dissimilarity to wild-type β-globin transcripts that rapidly disappear from the site of transcription below a detection threshold upon actinomycin D treatment (Figure 3), the 5′ splice site βSM and 3′ end formation βIVSI processing mutant RNA molecules are retained near the gene locus. Importantly, after 30 min of actinomycin D treatment the number of MELβIVSI or MELβSM cells with RNA foci in the nucleus was reduced by 50% and after 1 h to <1% (data not shown). This may reflect either degradation of the arrested RNAs or a protracted release from the site of transcription.

Since actinomycin D may induce pleiotropic effects on cells, we also assessed the effect of α-amanitin on the kinetics of human β-globin RNA release from the site of transcription. α-Amanitin is a cyclic peptide which binds with high affinity to the large subunit of RNA polymerase II (Cochet-Meilhac and Chambon, 1974; Lutter, 1982) thereby inhibiting transcription (Kedinger *et al.*, 1970; Lindell *et al.*, 1970). Unlike actinomycin D, α-amanitin penetrates slowly into cultured cells and requires a number of hours to inhibit transcription *in vivo* (Nguyen *et al.*, 1996). We therefore performed a time course analysis of the effect of α-amanitin on MEL cells transfected with the wild-type human β-globin gene (Figure 5B, MELβWT). The results show that after 1 h of treatment the proportion of cells with intranuclear foci remains essentially unaltered, whereas significant decreases are observed following exposure to the drug for 2, 2.5 and 3 h.

When a similar analysis was performed on MEL cells transfected with either the 5′ splice site mutant (MELβSM) or the mutant lacking IVS-II and defective in 3′ end formation (MELβIVSI), the proportion of cells with visible nuclear foci remained unaltered during the first 2 h of α-amanitin treatment (Figure 5B, MELβSM and MELβIVSI). However, after 2.5 h there is a significant decrease in the proportion of labelled MELβIVSI cells, whereas no change is detected in MELβSM cells. After 3 h of treatment the proportion of labelled MELβIVSI cells remains unchanged, while a significant decrease is detected for the first time in MELβSM cells. Therefore, the disappearance of the intranuclear signal of β-globin RNA induced by α-amanitin occurs with significantly slower kinetics in cells transfected with the mutant βSM and βIVSI gene constructs than in those cells harbouring the βWT transgene. Similarly, treatment of MELβWT, MELβSM and MELβIVSI cells with the purine nucleoside analogue DRB, a specific inhibitor of processive transcrip-

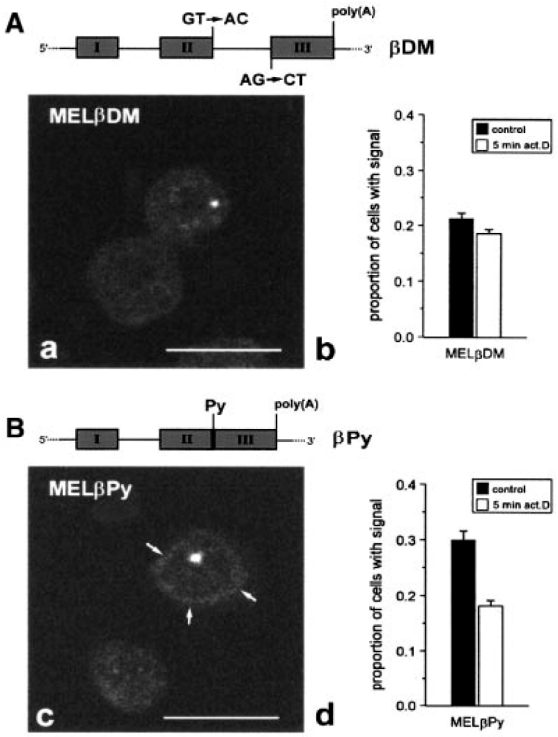


Fig. 6. Retention at the site of transcription correlates with the ability of β-globin RNA mutants to support at least partial spliceosome assembly. (**A**) (**a**) MEL cells transfected with the βDM construct which contains a 5′ splice donor (GT→AC) and a 3′ splice acceptor (AG→CT) double mutation of IVS-II. These MELβDM cells were induced to undergo erythroid differentiation for 4 days and hybridized with the full-length RNA probe. Note: a focal signal is readily visible in the nucleus, but no cytoplasmic staining is detected. (**b**) Quantitative analysis of the effect of actinomycin D on RNA release. Cells were induced for 2 days and hybridized with the full-length RNA probe, as described in the legend to Figure 5c. The results show that actinomycin D does not significantly affect the mean proportion of cells with signal in the nucleus [Student's *t*-test, $t_{(58)} = 2.04$, $p = 0.05$]. (**B**) (**c**) MELβPy cells harbour the βPy construct, which bears a 21 bp polypyrimidine tract substitution of IVS-II. After 4 days of erythroid differentiation these cells were hybridized with the full-length RNA probe. Note the presence of an intranuclear focus with additional staining of the cytoplasm (arrows). (**d**) Quantitative analysis shows that actinomycin D induces a highly significant decrease in the mean proportion of MELβPy cells with visible intranuclear foci [Student's *t*-test, $t_{(48.6)} = 6.24$, $p = 0.0001$]. Cells were either fixed in formaldehyde and permeabilized with Triton X-100/saponin (a and c) or fixed in formaldehyde/acetic acid and digested with pepsin (b and d).

tion that blocks phosphorylation of the CTD of RNA polymerase II (see Bentley, 1995), resulted in a significantly faster disappearance of intranuclear βWT RNA signal compared with that observed with the two mutants (data not shown).

In order to gain further insight into the mechanism of retention of transcripts at the site of transcription caused by defects in RNA processing, experiments were conducted with MELβDM and MELβPy cells. The 'double mutant' βDM construct contains a 5′ (GT→AC) and 3′ (AG→CT) splice site mutation (Figure 6A, upper panel). In the βPy construct, the second intron was removed and replaced by a 21 bp sequence corresponding to the polypyrimidine (Py) tract of IVS-II and devoid of all other splicing signals (Figure 6B, upper panel). Biochemical analysis has shown that βDM and βPy transcripts are, as expected, not spliced at the IVS-II position but are still able to undergo 3′ end formation at a level ~30% of that observed with the wild-type βWT gene (F.Geraghty and M.Antoniou, in preparation). In addition, whereas very low levels of unspliced, double splice site mutant βDM RNA reach the cytoplasm, βPy (Py tract/IVS-II substitution) transcripts are transported to the cytoplasm at 30% of βWT (F.Geraghty and M.Antoniou, in preparation).

Upon induced erythroid differentiation for 4 days, the full-length RNA probe (Figure 1A) produces focal hybridization signals in the nuclei of MEL cells transfected with both the double splice site βDM (MELβDM; Figure 6A, a) and the Py tract/IVS-II substitution (MELβPy; Figure 6B, c) mutant constructs. No cytoplasmic staining is detected in the MELβDM cells (Figure 6A, a), whereas a faint labelling can be observed with MELβPy cells (Figure 6B, c, arrows). Following treatment with actinomycin D for 5 min, the proportion of MELβDM cells showing a nuclear signal for human β-globin RNA is not significantly altered (Figure 6A, b). In contrast, the same treatment induces a highly significant decrease in the number of MELβPy cells that score positive for a focal signal of human β-globin RNA in the nucleus (Figure 6B, d).

These data indicate that the majority of double splice site βDM transcripts are retained at the site of transcription whereas newly synthesized Py tract/IVS-II substitution mutant βPy RNA molecules are partially released and eventually reach the cytoplasm. As the first intron is correctly spliced in all mutants analysed (Collis *et al.*, 1990; Antoniou *et al.*, 1998 and data not shown) and the efficiency of 3′ end formation of βDM and βPy mutant RNA is 30% of wild-type in both cases (F.Geraghty and M.Antoniou, in preparation), this suggests that the splicing factors that are still capable of binding to the mutant second intron of βDM but fail to interact with the βPy variant are contributing to retention.

Discussion

The βLCR/MEL cell system was chosen for these studies as it affords high physiological levels of gene expression from within a natural chromatin context (Blom van Assendelft *et al.*, 1989; Talbot *et al.*, 1989; Collis *et al.*, 1990). We therefore selected clones or pools of stably transfected MEL cells with a high transgene copy number in order to maximize the sensitivity of detecting human β-globin transcripts at all stages of their synthesis, maturation and transport within the nucleus. Interestingly, despite these advantages and the use of hybridization conditions which allow virtually 100% access of the probe to the target RNA within the nucleus (Wijgerde *et al.*, 1995), we were only able to detect human β-globin RNA as a focal concentration near the site of transcription (Figure 1). In addition, the use of oligonucleotide probes that span the intron–exon boundaries of the human β-globin gene also show that the first intron is spliced while the transcript is still at the gene locus, whereas removal of the second intron takes place either immediately prior to rapid release

from the site of transcription or at some other location within the nucleus after transport from the region of the gene template (Figure 2). The failure to visualize human β-globin RNA transcripts at a location other than in the vicinity of the gene locus clearly reflects the sensitivity limits of the experimental procedure that is unable to detect single RNA molecules in transit through the nucleoplasm. These results also imply that the mature mRNA does not concentrate in any other nuclear compartment as it is being transported to the cytoplasm and therefore in all likelihood follows a broad diffuse pathway as has been described for rat β-actin gene transcripts (Femino *et al.*, 1998).

The principal discovery of this study was the demonstration that pre-mRNA mutants unable to be exported to the cytoplasm due to the inability to undergo splicing or 3′ end processing, are retained within the nucleus in close proximity to the site of transcription (Figures 4–6). Our experimental approach was based on the direct *in situ* visualization of wild-type and mutant β-globin RNAs in the nucleus of cells treated with the transcription inhibitors actinomycin D, α-amanitin and DRB. A similar type of analysis utilizing transcription inhibitors in conjunction with RNA *in situ* hybridization has recently been used to determine the lifetime of α-, β- and γ-globin primary transcripts in fetal liver cells obtained from transgenic mice (Gribnau *et al.*, 1998). These data reveal that with probes complementary to the 5′ end of human β-globin primary transcripts, focal signals are still visible after 15 min of DRB treatment. This is consistent with previous studies indicating that DRB does not affect initiation of transcription but aborts elongating transcripts ~400–600 bp from the initiation site. Interestingly, the signal produced by 3′ end probes had completely disappeared after 7.5 min of DRB treatment. Therefore, the kinetics of release of primary transcripts in fetal liver cells from the site of transcription in single-copy human β-globin transgenic mice appears very similar to that observed in multicopy transfected MEL cells.

Importantly, our results also indicate that in order to be informative on the kinetics of release of RNA from the vicinity of the gene locus, exposure to transcription inhibitors should be for short periods. After prolonged drug treatments no difference is observed between wild-type and mutant RNAs, presumably due to degradation of the arrested transcripts.

One possible explanation for the retention of human β-globin pre-mRNA processing mutants in the vicinity of the site of transcription is that the release of transcripts is blocked by the stalled processing machinery attached to the nuclear matrix (Verheijen *et al.*, 1986; Smith *et al.*, 1989; Blencowe *et al.*, 1994). As spliceosome assembly and splicing are normally cotranscriptional events, it is expected that spliceosomes attach to the nuclear matrix at the site of transcription. Therefore, stalling of a mutant pre-mRNA due to impaired splicing should occur in the vicinity of the site of transcription. The finding that both splicing and polyadenylation factors can associate with the CTD of RNA polymerase II (Mortillaro *et al.*, 1996; Yuryev *et al.*, 1996; Du and Waren, 1997; McCracken *et al.*, 1997; reviewed by Steinmetz, 1997), also implies that RNA processing mutants may be retained close to the DNA template by remaining tethered to a stalled or abnormal processing machinery associated with the CTD of the polymerase. The range of pre-mRNA mutants we have analysed further suggests that assembly of the processing machinery for both splicing and 3′ end formation is involved in the retention mechanism.

Importantly, it has been reported that a hyperphosphorylated form of the large subunit of RNA polymerase II associates with the nuclear matrix (Mortillaro *et al.*, 1996; Vincent *et al.*, 1996). It is therefore conceivable that RNA polymerase II may interact directly with the nuclear matrix via the phosphorylated CTD. Alternatively, RNA polymerase II may be indirectly associated with the nuclear matrix through the association of RNA processing components with both the matrix (Verheijen *et al.*, 1986; Smith *et al.*, 1989; Blencowe *et al.*, 1994) and the CTD (see Steinmetz, 1997). In either case, retention of pre-mRNA processing mutants near the site of transcription could be explained by stalled processing machinery that is tethered to the nuclear matrix via the CTD.

A prediction of the model in which both the splicing and the 3′ end processing machinery are associated with the CTD of RNA polymerase II (Steinmetz, 1997) is that a normal mRNA would be released from the CTD as it is processed, whereas a pre-mRNA processing mutant would remain bound to the processing machinery and therefore to the CTD. Consequently, the polymerase may not be released from the gene at the termination of transcription of a mutated RNA. This would result in oncoming RNA polymerase molecules stalling on the template and therefore reducing the overall rate of synthesis. Furthermore, if RNA processing mutants fail to be released from the site of transcription, this should result in a local accumulation and consequent increase in intensity of the foci produced by *in situ* hybridization. However, a careful examination of the wild-type MELβWT and 5′ splice site mutant MELβSM cells that harbour a similar transgene copy number (data not shown), shows that there is no increase in signal intensity (compare Figures 1B, d and 4A, a). This observation is consistent with previously reported biochemical analysis indicating that in transfected cells the steady-state levels of this mutant RNA within the nucleus is the same as that observed with the wild-type normal transcript (Antoniou *et al.*, 1998). This is consistent with the notion that mutated RNAs are stalled on the template with a concurrent feedback mechanism to the transcription machinery. Without such a feedback the cells with mutant genes should have accumulated a significantly higher amount of RNA at the site of transcription than those containing the normal β-globin gene.

Studies performed on Balbiani ring pre-mRNAs expressed in the salivary gland cells of *Chironomus tentans*, suggest that splicing may occur either during or after transcription, depending on the position of the intron in the gene. In this particular system, introns located near the 5′ end of the gene are excised cotranscriptionally while introns closer to the 3′ end are more frequently spliced after release of the RNA into the nucleoplasm (Bauren and Wieslander, 1994; Wetterberg *et al.*, 1996). An immediate question raised by these results is whether introns spliced post-transcriptionally assemble the spliceosome at the site of transcription or after release of the pre-mRNA into the nucleoplasm. If the spliceosome assembles cotranscriptionally and excision of the intron

can occur in the nucleoplasm, this would imply that completion of the splicing reaction is not required for release from the site of transcription. In fact, as the assembly of a functional spliceosome involves a dynamic and timely rearrangement of its components (reviewed by Madhani and Guthrie, 1994; Ares and Weiser, 1995), it is feasible to anticipate that splicing factors may detach from the CTD prior to the final catalytic steps of splicing. On the other hand, a post-transcriptional assembly of the spliceosome would argue that interaction of splicing factors with the CTD is not essential for loading the spliceosome on a pre-mRNA.

A striking exception to the rule that mRNAs can be exported only after completion of processing occurs in retroviruses, which have evolved a mechanism that allows the nuclear export of unspliced forms of viral RNAs. This mechanism is best characterized in human immunodeficiency virus type 1 (HIV-1) and involves the virally encoded protein Rev (for a recent review see Stutz and Rosbash, 1998). Upon binding of Rev to the Rev response element (RRE) present in the intron of immature viral mRNA, the complex is transported to the cytoplasm by virtue of interacting with CRM1/exportin 1 through a leucine-rich nuclear export signal present at the Rev C-terminal end. Thus, the association of Rev with an RRE promotes the interaction of the RRE-containing mRNA with exportin and consequently its export from the nucleus. In addition, it is possible that binding of Rev to the intronic RRE interferes with spliceosome assembly, thereby contributing to its premature release from the processing machinery. Consistent with this idea, there is evidence that Rev specifically blocks assembly of U4/U6 and U5 snRNPs into the spliceosome (Kjems and Sharp, 1993). Our observations indicating that unprocessed β-globin RNAs are retained at the site of transcription suggest that Rev may have a dual function in promoting export of unspliced viral RNAs. In addition to its well-established interaction with the exportin pathway, Rev may contribute directly to release of unspliced RNA from the spliceosome and hence from the site of transcription. Clearly, further experiments are needed to determine the spatial and temporal relationships between sites of transcription, spliceosome assembly and splicing in the nucleus.

Irrespective of the mechanism(s) responsible for the observed retention of mutant pre-mRNAs, a major conclusion from this study is that mechanisms which prevent export of pre-mRNA processing mutants to the cytoplasm, operate in close proximity to the site of transcription. Therefore, the efficiency of splicing and 3′ end formation appears to be rate limiting for the release of mRNAs pre-assembled with processing factors at the site of transcription.

Materials and methods

Gene constructs

The wild-type (βWT) and mutant βSM, βDM, βPy and βIVSI human β-globin genes were cloned in the microlocus LCR expression vector (Collis *et al.*, 1990), and are described in detail elsewhere (Antoniou *et al.*, 1998). Briefly, a β-globin gene harbouring a fully functional 89 bp deletion mutant of the second intron was used as the starting point for generating the βSM, βDM and βPy constructs (Antoniou *et al.*, 1998). The βVSI gene is as described previously (Collis *et al.*, 1990). The βWT, βSM, βDM and βPy genes extend to +1800 bp past the poly(A)-addition site, whereas the βIVSI construct terminates at +45 bp. This difference in the extent of 3′ flanking sequences does not in itself compromise the efficiency of 3′ end formation (Antoniou *et al.*, 1998). All these genes begin at a *Sna*BI site at –265 bp from the transcriptional start point and were cloned between the *Cla*I and *Asp*718 sites of the microlocus LCR expression vector (Collis *et al.*, 1990).

MEL cell transfections

The generation, maintenance and induced differentiation of stably transfected, G418 resistant MEL cell clones was as described previously (Antoniou, 1991). The MELβWT and MELβSM clones harbour ~14 copies of the transgene as a tandem array whereas MELβDM, MELβPy and MELβIVSI are large populations of stably transfected cells with an average transgene copy number of five (data not shown). Transgenes were confirmed to have integrated at a single chromosome site by fluorescence *in situ* hybridization of cells in metaphase (data not shown). Immunofluorescence was performed using an antibody specific for human globin (Immuno-rx, Augusta, GA, USA), as described (Fraser *et al.*, 1993). Actinomycin D (5 μg/ml), α-amanitin (100 μg/ml) and DRB (75 μM) were added to cells that had been induced to undergo erythroid differentiation for 2 days.

Probes used for in situ hybridization

Genomic cloned probes (see Figure 1A) were labelled with either digoxigenin-11-dUTP (Boehringer Mannheim) or dinitrophenyl-11-dUTP (DNP; Molecular Probes) by nick-translation (Lichter *et al.*, 1991). The full-length RNA probe extends over the entire transcribed region of the human β-globin gene and consists of a 3.7 kb fragment extending from the *Sna*BI site at –265 bp from the transcriptional start point to a *Bgl*II site at +1816 bp past the poly(A)-addition site. The 3′ RNA probe is a 771 bp fragment that extends from a position 212 bp upstream and 559 bp downstream of the human β-globin gene poly(A)-addition site. Therefore, the 3′ RNA probe by *in situ* hybridization detects nascent transcripts that have been transcribed past 212 nucleotides upstream of the polyadenylation site as well as those that have undergone termination and 3′ cleavage. As a result both 'terminating/nearly terminated' and 'terminated' transcripts will be detected. The DNA probe is the LCR expression vector (Collis *et al.*, 1990) into which the human β-globin genes under analysis were cloned.

Splice junction (SJ) oligonucleotide probes were purchased from Cruachem (UK):

Exon I–II spanning, SJ I/II, 5′-ACCACCAGCAGC/CTGCCCAGG-GCC-3′;
Exon II–III spanning, SJ II/III, 5′-GTTGCCCAGGAG/CCTGAAGTT-CTC-3′.

The forward slash mark indicates the exon boundaries. In addition to these 24 nucleotides complementary to the human β-globin sequence, the following stretch of non-specific sequence was added to increase the intensity of the hybridization signal and therefore sensitivity of the assay (Zhang *et al.*, 1994):

5′ end, 5′-TTTTTGCTTGCTTGCTT-3′;
3′ end, 5′-TTGCTTGCTTGCTT-3′.

The underlined bases show the positions of the nucleotides bearing an adjunct of DNP.

In situ hybridization

Cells were allowed to adhere onto poly-L-lysine coated coverslips and washed with phosphate buffered saline (PBS). The cells were then fixed with either 3.7% formaldehyde in PBS for 10 min, and permeabilized in 0.5% Triton X-100, 0.5% saponin (Zirbel *et al.*, 1993), or in 4% formaldehyde/5% acetic acid/0.9% NaCl and digested with 0.01% pepsin in 0.01 M HCl (Wijgerde *et al.*, 1995). Cloned probes were hybridized for 16 h at 37°C in 50% formamide/2× SSC/10% dextran sulfate/50 mM sodium phosphate pH 7.0. Post-hybridization washes were in 50% formamide/2× SSC (3× 5 min at 45°C) and either 2× SSC (3× 5 min at 45°C) for the full-length and 3′ RNA probes, or 0.5× SSC (3× 5 min at 45°C) for the DNA probe. Hybridization with oligonucleotides was performed in 20% formamide/2× SSC/10% dextran sulphate/0.2% BSA/1 μg/μl tRNA, at 37°C for 3 h. Post-hybridization washes were in 20% formamide/2× SSC (3× 5 min at 42°C) and 2× SSC (3× 5 min at 42°C). The sites of hybridization were visualized using antibodies directed against either digoxigenin (Boehringer Mannheim) or DNP (Molecular Probes) and appropriate secondary antibodies coupled to fluorescein, rhodamine or Texas Red (Vector Laboratories; Jackson

ImmunoResearch). The staining of total DNA was performed after *in situ* hybridization by incubating cells for 10 min with 0.5 μM TO-PRO-3 (Molecular Probes).

Double hybridization experiments were performed sequentially. After the first hybridization and detection steps as described above, cells were fixed with 3.7% formaldehyde in PBS for 10 min and then hybridized again. As controls for the double-labelling experiments, the complete double-hybridization procedure was carried out omitting either the DNA or the RNA probe. Under these conditions no DNA or RNA signal was detected, respectively, confirming the specificity of each labelling reaction. In addition, cells were hybridized with both DNA and RNA probes under non-denaturing conditions. In these experiments the DNA probe produces a very faint fluorescent signal in some nuclei (data not shown). These faint signals produced by the DNA probe and which colocalize with the foci produced by hybridization with the RNA probe, are likely to represent a combination of transcripts from the neomycin-resistance (TK *neo*^{r}) gene and those arising from within the LCR (Collis *et al.*, 1990; Ashe *et al.*, 1997) which are present on the plasmid expression cassette. Digestion with RNase A before hybridization completely abolished labelling (data not shown), confirming that the observed signals correspond to RNA hybridization. Moreover, no labelling was observed in untransfected MEL cells after 4 days of erythroid differentiation, indicating that the hybridization signal is specific for human globin RNA (data not shown).

Microscopy

Samples were examined with a Zeiss LSM 410 microscope. Confocal microscopy was performed using argon ion (488 nm) and HeNe (543 nm) lasers to excite FITC and TxRed/rhodamine fluorescence, respectively.

Acknowledgements

The excellent technical support of Jacky Hurst is gratefully acknowledged. This study was supported by grants from the Junta Nacional de Investigaçao Cientıfica e Tecnologica /Programme PRAXIS XXI, Portugal, the Medical Research Council, UK, the Human Capital and Mobility Programme and Biomed 2 programmes of the European Union. N.C. and S.P. were supported by PRAXIS XXI fellowships.

References

Antoniou,M. (1991) Induction of erythroid-specific expression in murine erythroleukemia (MEL) cells. In Murray,E.J. (ed.), *Methods in Molecular Biology vol. 7: Gene Transfer and Expression Protocols*. The Humana Press Inc., Clifton, NJ, pp. 421–434.

Antoniou,M., Carmo-Fonseca,M., Ferreira,J. and Lamond,A.I.J. (1993) Nuclear organisation of splicing snRNPs during differentiation of murine erythroleukemia cells *in vitro*. *J. Cell Biol.*, **123**, 1055–1068.

Antoniou,M., Geraghty,F., Hurst,J. and Grosveld,F. (1998) Efficient 3′-end formation of human β-globin mRNA *in vivo* requires sequences within the last intron but occurs independently of the splicing reaction. *Nucleic Acids Res.*, **26**, 721–729.

Ares,M. and Weiser,B. (1995) Rearrangement of snRNA structure during assembly and function of the spliceosome. *Prog. Nucleic Acid Res. Mol. Biol.*, **50**, 131–159.

Ashe,H.L., Monks,J., Wijgerde,M., Fraser,P. and Proudfoot,N.J. (1997) Intergenic transcription and transinduction of the human β-globin locus. *Genes Dev.*, **11**, 2494–2509.

Bauren,G. and Wieslander,L. (1994) Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell*, **76**, 183–192.

Bauren,G., Jiang,W.-Q., Bernholm,K., Gu,F. and Wieslander,L. (1996) Demonstration of a dynamic, transcription-dependent organization of pre-mRNA splicing factors in polytene nuclei. *J. Cell Biol.*, **133**, 929–942.

Bentley,D.L. (1995) Regulation of transcriptional elongation by RNA polymerase II. *Curr. Opin. Genet. Dev.*, **5**, 210–216.

Beyer,A.L. and Osheim,Y.N. (1988) Splice site selection, rate of splicing and alternative splicing on nascent transcripts. *Genes Dev.*, **2**, 754–765.

Blencowe,B.J., Nickerson,J.A., Issner,R., Penman,S. and Sharp,P.A. (1994) Association of nuclear matrix antigens with exon-containing splicing complexes. *J. Cell Biol.*, **127**, 593–607.

Blom van Assendelft,G., Hanscombe,O., Grosveld,F. and Greaves,D.R. (1989) The β-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell*, **56**, 969–977.

Chang,D.D. and Sharp,P.A. (1989) Regulation by HIV Rev depends upon recognition of splice sites. *Cell*, **59**, 789–795.

Cochet-Meilhac,M. and Chambon,P. (1974) Animal DNA-dependent RNA polymerases. 11. Mechanism of the inhibition of RNA polymerases B by amatoxins. *Biochim. Biophys. Acta*, **353**, 160–184.

Collis,P., Antoniou,M. and Grosveld,F. (1990) Definition of the minimal requirements within the human β-globin gene and the dominant control region for high level expression. *EMBO J.*, **9**, 233–240.

Darnell,J.E., Philipson,L., Wall,R. and Adesnik,M. (1971) Polyadenylic acid sequences: role in conversion of nuclear RNA into messenger RNA. *Science*, **174**, 507–510.

Du,L. and Warren,S. (1997) A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing. *J. Cell Biol.*, **136**, 5–18.

Eckner,R., Ellmeier,W. and Birnstiel,M.L. (1991) Mature mRNA 3′ end formation stimulates RNA export from the nucleus. *EMBO J.*, **10**, 3513–3522.

Elliot,D.J., Stutz,F., Lescure,A. and Rosbash,M. (1994) mRNA export. *Curr. Opin. Genet. Dev.*, **4**, 305–309.

Femino,A.M., Fay,F.S., Fogarty,K. and Singer,R.H. (1998) Visualisation of single RNA transcripts *in situ*. *Science*, **280**, 585–590.

Fraser,P., Pruzina,S., Antoniou,M. and Grosveld,F. (1993) Each hypersensitive site of the human β-globin locus control region confers a different developmental pattern of expression on the globin genes. *Genes Dev.*, **7**, 106–113.

Goldberg,I.H., Rabinowitz,M. and Reich,E. (1962) Basis of actinomycin action, 1. DNA binding and inhibition of RNA-polymerase synthetic reactions by actinomycin. *Proc. Natl Acad. Sci. USA*, **48**, 2094–2101.

Gribnau,J., Boer,E., Trimborn,T., Wijgerde,M., Milot,E., Grosveld,F. and Fraser,P. (1998) Chromatin interaction mechanism of transcriptional control *in vivo*. *EMBO J.*, **17**, 6020–6027.

Hamm,J. and Mattaj,I.W. (1990) Monomethylated cap structures facilitate RNA export from the nucleus. *Cell*, **63**, 109–118.

Huang,S. and Spector,D.L. (1996) Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription. *J. Cell Biol.*, **133**, 719–732.

Izaurralde,E. and Mattaj,I.W. (1992) Transport of RNA between nucleus and cytoplasm. *Semin. Cell Biol.*, **3**, 279–288.

Jarmolowski,A., Boelens,W.C., Izaurralde,E. and Mattaj,I.W. (1994) Nuclear export of different classes of RNA is mediated by specific factors. *J. Cell Biol.*, **124**, 627–635.

Kedinger,C., Gniazdowski,M., Mandel,J.L.,Jr, Gissinger,F. and Chambon,P. (1970) Alpha-amanitin: a specific inhibitor of one of two DNA-pendent RNA polymerase activities from calf thymus. *Biochem. Biophys. Res. Commun.*, **38**, 165–171.

Kersten,W., Kersten,H. and Rauen,H.M. (1960) Action of nucleic acids on the inhibition of growth by actinomycin of *Neurospora crassa*. *Nature*, **187**, 60–61.

Kjems,J. and Sharp,P.A. (1993) The basic domain of Rev from human immunodeficiency virus type 1 specifically blocks the entry of U4/U6 and U5 small nuclear ribonucleoprotein in spliceosome assembly. *J. Virol.*, **67**, 4769–4776.

Kramer,A. (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu. Rev. Biochem.*, **65**, 367–409.

Lamond,A.I., Konarska,M.M. and Sharp,P. (1987) A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation. *Genes Dev.*, **1**, 532–543.

Lee,M.S. and Silver,P.A. (1997) RNA movement between the nucleus and the cytoplasm. *Curr. Opin. Genet. Dev.*, **7**, 212–219.

Legrain,P. and Rosbash,M. (1989) Some *cis*- and *trans*-acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell*, **57**, 573–583.

LeMaire,M.F. and Thummel,C.S. (1990) Splicing precedes poly-adenylation during *Drosophila* E74A transcription. *Mol. Cell. Biol.*, **10**, 6059–6063.

Lichter,P., Boyle,A.L., Cremer,T. and Ward,D.C. (1991) Analysis of genes and chromosomes by non-isotopic *in situ* hybridisation. *Genet. Anal. Tech. Appl.*, **8**, 24–35.

Lindell,T.J., Weinberg,F., Morris,P.W., Roeder,R.G. and Rutter,W.J. (1970) Specific inhibition of nuclear RNA polymerase II by α-amanitin. *Science*, **170**, 447–449.

Lutter,L.C. (1982) Photoreactivation of amanitin-inhibited RNA polymerase II. *J. Biol. Chem.*, **257**, 1577–1578.

Madhani,H.D. and Guthrie,C. (1994) Dynamic RNA–RNA interactions in the spliceosome. *Annu. Rev. Genet.*, **28**, 1–26.

McCracken,S., Fong,N., Yankulov,K., Ballantyne,S., Pan,G., Greenblatt,J., Patterson,S., Wickens,M. and Bentley,D.L. (1997) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, **385**, 357–361.

Milot,E. *et al.* (1996) Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell*, **87**, 105–114.

Moore,M.J., Query,C.C. and Sharp,P. (1993) Splicing of precursors to mRNA by the spliceosome. In Gesteland,R. and Atkins,J. (eds), *The RNA World.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.

Mortillaro,M.J., Blencowe,B.J., Wei,X., Nakayasu,H., Du,L., Warren, S.L., Sharp,P.A. and Berezney,R. (1996) A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc. Natl Acad. Sci. USA*, **93**, 8253–8257.

Nakielny,S., Fischer,U., Michael,W.M. and Dreyfuss,G. (1997) RNA transport. *Annu. Rev. Neurosci.*, **30**, 269–301.

Neugebauer,K.M. and Roth,M.B. (1997) Distribution of pre-mRNA splicing factors at sites of RNA polymerase II transcription. *Genes Dev.*, **11**, 1148–1159.

Nguyen,V.T., Giannoni,F., Dubois,M.F., Seo,S.J., Vigneron,M., Kedinger,C. and Bensaude,O. (1996) *In vivo* degradation of RNA polymerase II largest subunit triggered by α-amanitin. *Nucleic Acids Res.*, **24**, 2924–2929.

Nigg,E.A. (1997) Nucleocytoplasmic transport: signals, mechanisms and regulation. *Nature*, **386**, 779–787.

SAS Institute Inc. (1990) *SAS User's Guide to Statistics*, *Version 6*. 4th edn. SAS Institute Inc., Cary, NC.

Smith,H.C., Harris,S.G., Zillmann,M. and Berget,S.M. (1989) Evidence that a nuclear matrix protein participates in pre-messenger RNA splicing. *Exp. Cell Res.*, **182**, 521–533.

Steinmetz,E.J. (1997) Pre-mRNA processing and the CTD of RNA polymerase II: the tail that wags the dog? *Cell*, **89**, 491–494.

Stutz,F. and Rosbash,M. (1998) Nuclear RNA export. *Genes Dev.*, **12**, 3303–3319.

Talbot,D., Collis,P., Antoniou,M., Vidal,M., Grosveld,F. and Greaves, D.R. (1989) A dominant control region from the human β-globin locus conferring integration site-independent gene expression. *Nature*, **338**, 352–355.

Tennyson,C., Klamut,H.J. and Worton,R.G. (1995) The human dystrophin gene requires 16 h to be transcribed and is cotranscriptionally spliced. *Nature Genet.*, **9**, 184–190.

Verheijen,R., Kuijpers,H., Vooijs,P., VanVenrooij,W. and Ramaekers,F. (1986) Distribution of the 70k U1 RNA-associated protein during interphase and mitosis. Correlation with other U RNP particles and proteins of the nuclear matrix. *J. Cell Sci.*, **86**, 173–190.

Vincent,M., Lauriault,P., Dubois,M.-F., Lavoie,S., Bensaude,O. and Chabot,B. (1996) The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes. *Nucleic Acids Res.*, **24**, 4649–4652.

Visa,N., Alzhanova-Ericsson,A.T., Sun,X., Kiseleva,E., Bjorkroth,B., Wurtz,T. and Daneholt,B. (1996a) A pre-mRNA-binding protein accompanies the RNA from the gene through the nuclear pores and into polysomes. *Cell*, **84**, 253–264.

Visa,N., Izaurralde,E., Ferreira,J., Daneholt,B. and Mattaj,I. (1996b) A nuclear cap binding complex binds Balbiani ring pre-mRNA co-transcriptionally and accompanies the ribonucleoprotein particle during nuclear export. *J. Cell Biol.*, **133**, 5–14.

Wetterberg,I., Bauren,G. and Wieslander,L. (1996) The intranuclear site of excision of each intron in Balbiani ring 3 pre-mRNA is influenced by the time remaining to transcription termination and different excision efficiencies for the various introns. *RNA*, **2**, 641–651.

Wijgerde,M., Grosveld,F. and Fraser,P. (1995) Transcription complex stability and chromatin dynamics *in vivo*. *Nature*, **377**, 209–213.

Wu,Z., Murphy,C., Callan,H.G. and Gall,J.G. (1991) Small nuclear ribonucleoproteins and heterogeneous nuclear ribonucleoproteins in the amphibian germinal vesicle: loops, spheres and snurpeosomes. *J. Cell Biol.*, **113**, 465–483.

Xing,Y., Johnson,C.V., Dobner,P.R. and Lawrence,J.B. (1993) Higher level organisation of individual gene transcription and RNA splicing. *Science*, **259**, 1326–1330.

Yuryev,A., Patturajan,M., Litingtung,Y., Joshi,R.V., Gentile,C., Gebara,M. and Corden,J.L. (1996) The CTD of RNA polymerase II interacts with a novel set of SR-like proteins. *Proc. Natl Acad. Sci. USA*, **93**, 6975–6980.

Zhang,G., Taneja,K.L., Singer,R.H. and Green,M.R. (1994) Localisation of pre-mRNA splicing in mammalian nuclei. *Nature*, **372**, 809–812.

Zhang,G., Zapp,M.L., Yan,G. and Green,M. (1996) Localisation of HIV-1 RNA in mammalian nuclei. *J. Cell Biol.*, **135**, 9–18.

Zirbel,R.M., Mathieu,U.R., Kurz,A., Cremer,T. and Lichter,P. (1993) Evidence for a nuclear compartment of transcription and splicing located at chromosome domain boundaries. *Chromosome Res.*, **1**, 93–106.

Received January 12, 1999; revised and accepted March 18, 1999

2. *In vivo* recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei

Having localised the retention of the abnormally processed *HBB* transcripts to the site of transcription, indicating that the quality control mechanism was probably co-transcriptional, the next step was to determine the molecular players involved in this mechanism. The retention could be directly mediated by proteins that were bound to the nascent transcripts and were stalled due to the incomplete processing or could be due to the absence of recruitment of proteins essential for the release and/or transport of the transcripts from the gene locus. To test the later hypothesis we decided to study the recruitment of exon junction complex (EJC) proteins and mRNA export factors to the retention sites in the nucleus. From this work we concluded that EJC proteins were recruited *in vivo* to sites of normal *HBB* transcription but not to the mutant *HBB* transcription sites. These results suggest that EJC proteins bind stably to the pre-mRNA co-transcriptionally and the accumulation of EJC proteins at the transcription sites requires efficient processing. No concentration of the export factors NXF1/TAP, p15 was detected on nascent transcripts, arguing that in mammalian cells these proteins bind the mRNA shortly before or after release from the sites of transcription

Work published in: RNA Vol.10 No.4 pp.622-633, 2004

I would like to stress that the results presented and discussed in this section were equally obtained by me and Dr. Célia Carvalho. I contributed with all the results produced using as a model system the murine erythroleukemia (MEL) cells stably transfected with the human β-globin gene and Dr. Célia Carvalho contributed with the results produced using HeLa cells infected with adenovirus. Inês Condado participated in some FISH experiments, Dr. Michael Antoniou generated the MEL cell clones analysed and Dr. Benjamin Blencowe gave a valuable contribution in the RNase protection assay.

In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei

NOÉLIA CUSTÓDIO,[1,4] CÉLIA CARVALHO,[1,4] INÊS CONDADO,[1] MICHAEL ANTONIOU,[2] BENJAMIN J. BLENCOWE,[3] and MARIA CARMO-FONSECA[1]

[1]Institute of Molecular Medicine, Faculty of Medicine, University of Lisbon, Lisbon 1649-028, Portugal
[2]Division of Medical and Molecular Genetics, GKT School of Medicine, Guy's Hospital, London SE1 9RT, UK
[3]Banting and Best Department of Medical Research, C.H. Best Institute, University of Toronto, Toronto, Ontario M5G 1L6, Canada

ABSTRACT

Studies over the past years indicate that there is extensive coupling between nuclear export of mRNA and pre-mRNA processing. Here, we visualized the distribution of exon junction complex (EJC) proteins and RNA export factors relative to sites of abundant pre-mRNA synthesis in the nucleus. We analyzed both HeLa cells infected with adenovirus and murine erythroleukemia (MEL) cells stably transfected with the human β-globin gene. Using in situ hybridization and confocal microscopy, we observe accumulation of EJC proteins (REF/Aly, Y14, SRm160, UAP56, RNPS1, and Magoh) and core spliceosome components (U snRNPs) at sites of transcription. This suggests that EJC proteins bind stably to pre-mRNA cotranscriptionally. No concentration of the export factors NXF1/TAP, p15, and Dbp5 was detected on nascent transcripts, arguing that in mammalian cells these proteins bind the mRNA shortly before or after release from the sites of transcription. These results also suggest that binding of EJC proteins to the mRNA is not sufficient to recruit TAP-p15, consistent with recent findings showing that the EJC does not play a crucial role in mRNA export. Contrasting to the results obtained in MEL cells expressing normal human β-globin transcripts, mutant pre-mRNAs defective in splicing and 3′end processing do not colocalize with SRm160, REF, UAP56, or Sm proteins. This shows that the accumulation of EJC proteins at transcription sites requires efficient processing of the nascent pre-mRNAs, arguing that transcription per se is not sufficient for the stable assembly of the EJC.

Keywords: pre-mRNA splicing; mRNA export; exon junction complex; NXF1/TAP

INTRODUCTION

In eukaryotes, messenger RNAs are transcribed in the nucleus as precursor forms (pre-mRNAs). Immediately upon synthesis, nascent transcripts associate with proteins forming ribonucleoprotein (RNP) particles, the protein content of which evolves throughout the lifetime of a mRNA (for a recent review, see Dreyfuss et al. 2002). The proteins associated with messenger ribonucleoprotein particles (mRNPs) play key roles in all aspects of the RNA metabolism, from nucleocytoplasmic transport to cytoplasmic localization, translational efficiency, and decay. In particular, a specific subset of proteins including SRm160, RNPS1, Y14, Mago, DEK, REF/Aly, and UAP56 form the exon junction complex (EJC), which associates with mRNAs as a consequence of splicing (Blencowe et al. 1998; Mayeda et al. 1999; Kataoka et al. 2000; Le Hir et al. 2000a,b; McGarvey et al. 2000; Zhou et al. 2000; Luo et al. 2001). SRm160 and RNPS1 were originally characterized as activators of pre-mRNA splicing (Fleckner et al. 1997; Blencowe et al. 1998; Mayeda et al. 1999). More recently, SRm160 was shown to promote transcript 3′-end cleavage (McCracken et al. 2002), whereas RNPS1 couples splicing to nonsense-mediated decay (Lykke-Andersen et al. 2001). Y14 is involved in mRNA quality control via the nonsense-mediated mRNA decay (NMD) pathway (Lykke-Andersen et al. 2001), and together with Mago, is required for the proper cytoplasmic localization of *oskar* mRNA during *Drosophila* development (Hachet and Ephrussi 2001; Mohr et al. 2001). Y14 and Mago form a tight heterodimer in vivo (Lau et al. 2003), and the association between the two proteins is essential for function in NMD (Fribourg et al. 2003). DEK has been involved in multiple functions ranging from splicing (McGarvey et al. 2000) to chromatin structure (Alexiadis et al. 2000) and transcriptional regulation (Faulkner et al. 2001). REF/Aly has been identified as a

[4]These authors contributed equally to this work.

Reprint request to: Maria Carmo-Fonseca, Institute of Molecular Medicine, Faculty of Medicine, Av. Prof. Egas Moniz, 1649-028 Lisboa, Portugal; e-mail: carmo.fonseca@fm.ul.pt; fax: 351-21-7951780.

Article and publication are at http://www.rnajournal.org/cgi/doi/10.1261/rna.5258504.

chaperone that regulates the activity of bZIP transcription factors (Virbasius et al. 1999), and also as a factor that facilitates the nuclear export of mRNA by interacting with the export receptor NXF1/TAP (for review, see Reed and Hurt 2002). UAP56 is a putative RNA helicase also implicated as a splicing factor required for early spliceosome assembly (Kistler and Guthrie 2001; Libri et al. 2001; Zhang and Green 2001). Contrasting with NXF1/TAP and UAP56, the depletion of which causes nuclear accumulation of mRNA, REF/Aly, RNPS1, SRm160, and Y14 are dispensable for bulk mRNA export (Gatfield and Izaurralde 2002; Longman et al. 2003; MacMorris et al. 2003). Accordingly, SR proteins were identified recently as adaptors that, in addition to REF/Aly, can mediate the interaction between NXF1/TAP and cellular mRNAs (Huang et al. 2003).

Translocation of mRNAs through the nuclear pore complex requires binding of an heterodimer composed of NXF1/TAP and NXT1/p15 (Mex67 and Mtr2 in *Saccharomyces cerevisiae*). NXF1/TAP, which interacts both with RNA-binding adapter proteins and components of the nuclear pore complex, is believed to be the major receptor for the export of mRNAs to the cytoplasm (for recent reviews, see Izaurralde 2002; Lei and Silver 2002b; Reed and Hurt 2002; Cullen 2003; Stutz and Izaurralde 2003). Another protein that interacts simultaneously with mRNPs and nucleoporins is Dbp5/Rat8 (Snay-Hodge et al. 1998; Tseng et al. 1998). Although in yeast Dbp5 is essential for mRNA export to the cytoplasm (Snay-Hodge et al. 1998; Tseng et al. 1998), depletion of the *Drosophila* homolog of Dbp5 does not result in the bulk nuclear retention of mRNAs (Gatfield et al. 2001).

How export-competent mRNPs assemble in the living cell nucleus and travel to the nuclear pores remains poorly understood. Here, we visualize the localization of EJC proteins and mRNA export factors in the nucleus of mammalian cells producing abundant pre-mRNA transcripts. First, we analyzed HeLa cells infected with adenovirus, because this virus recruits the host transcription and processing machinery to the sites of viral pre-mRNA synthesis (Jimenez-Garcia and Spector 1993; Pombo et al. 1994). In addition, we made use of murine erythroleukemia (MEL) cells stably transfected with the human β-globin gene under the control of the locus control region (LCR). The human β-globin gene, which integrates in the host cell genome as a tandem array, expresses at physiological levels upon induction of MEL cells to undergo terminal erythroid differentiation (Collis et al. 1990). Using in situ hybridization and confocal microscopy, we observe colocalization of EJC proteins (REF/Aly, Y14, SRm160, UAP56, RNPS1, and Magoh) and core spliceosome components (U snRNPs) at sites in the nucleus containing nascent adenoviral and β-globin transcripts. In contrast, the export factors NXF1/TAP, p15, and Dbp5 are not detected at sites of transcription. The results further show that REF/Aly, SRm160, UAP56, and spliceosomal snRNPs fail to accumulate at the site of transcription of a processing-defective β-globin mutant. Taken together, these data suggest that EJC proteins bind cotranscriptionally to mRNPs, and that efficient pre-mRNA processing is required for assembly of EJC proteins onto nascent transcripts in vivo.

RESULTS

Components of the EJC, but not NXF1/TAP, p15, or Dbp5, are recruited to sites of adenoviral transcription in the nucleus of infected HeLa cells

Adenovirus type 2 (Ad2) causes a productive infection of HeLa cells that proceeds through an infectious cycle of ~36 h. This cycle is conventionally divided into early and late stages, separated by the onset of viral DNA replication (for review, see Philipson et al. 1975). Adenoviruses enter the host cells by receptor-mediated endocytosis, penetrate the cytoplasm from endosomes, and deliver their DNA genome into the nucleus (Greber et al. 1993). Upon entry in the nucleus, transcription of the viral genome starts immediately, and the viral mRNAs transcribed at this early stage direct the synthesis of a small number of proteins that promote viral DNA replication. After the onset of viral DNA replication, which occurs at ~8 h post-infection, the remaining viral genomic information is expressed, yielding large quantities of the structural proteins that will eventually constitute new virus particles (Philipson et al. 1975; Flint 1986). During this period, the normal nuclear architecture undergoes a series of progressive changes (for review, see Pombo and Carmo-Fonseca 1995). The replication of Ad DNA produces single-stranded intermediates (ssDNA) that accumulate, forming viral inclusions in the nucleoplasm. These inclusions are readily identified using antibodies against the viral ssDNA-binding protein DBP. Following incorporation of a modified uridine-analog, it is possible to visualize by fluorescence microscopy the sites of RNA synthesis in the nucleus (Pombo et al. 1994). In noninfected cells, newly synthesized RNAs incorporating bromouridine are detected widespread throughout the nucleoplasm (Fig. 1A). In infected cells, nascent transcripts form ring-like structures (Fig. 1B) that surround the inclusions containing viral ssDNA (Fig. 1C). Previous work has indicated that these ring-like structures represent nascent viral mRNA transcripts (Pombo et al. 1994).

Consistent with previous evidence for cotranscriptional processing of pre-mRNA obtained in noninfected cells (Beyer and Osheim 1988; LeMaire and Thummel 1990; Wu et al. 1991; Xing et al. 1993; Bauren and Wieslander 1994; Zhang et al. 1994; Tennyson et al. 1995; Bauren et al. 1996; Huang and Spector 1996; Neugebauer and Roth 1997), in situ evidence suggests that splicing of Ad mRNAs occurs cotranscriptionally. Namely, spliceosome components including snRNPs and non-snRNP splicing proteins are recruited to the sites of viral mRNA synthesis (Pombo et al.

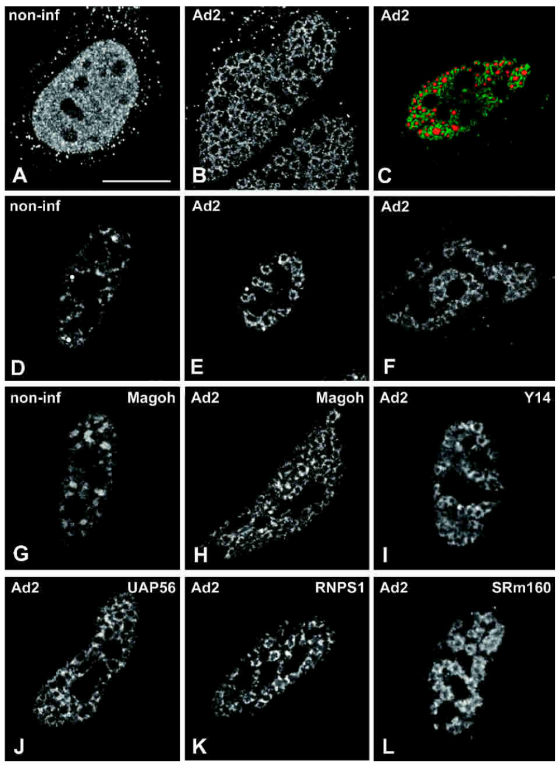


FIGURE 1. Localization of EJC proteins in adenovirus-infected cells. HeLa cells were either noninfected (*A*,*D*,*G*; non-inf) or infected with adenovirus for 14–16 h (*B*,*C*,*E*,*F*,*H*–*L*; Ad2). To detect nascent transcripts, cells were incubated with Br-U for 1 h (*A*–*C*). (*C*) A cell double-labeled for nascent transcripts (red) and viral ssDNA inclusions using anti-DBP antibody (green staining). To visualize spliceosome snRNPs, cells were immunolabeled with Y12 antibody directed against Sm proteins (*D*,*E*). Spliced viral mRNAs were detected by in situ hybridization using a probe (SJ1), which hybridizes with all mRNAs encoded by the highly expressed adenovirus major late transcription unit (Bridge et al. 1996) (*F*). The localization of EJC proteins was determined in cells expressing GFP–Magoh (*G*,*H*), zzY14 (*I*), GFP–UAP56 (*J*), GFP–RNPS1 (*K*). The distribution of SRm160 was visualized with specific rabbit polyclonal antibodies (*L*). Bar, 10 μm.

1994; Bridge et al. 1995; Gama-Carvalho et al. 1997, 2003). Furthermore, spliced major late-viral mRNAs are detected in the ring-like structures formed by the nascent transcripts (Bridge et al. 1996; Gama-Carvalho et al. 2003, Fig. 1D–F).

Here we analyzed whether protein components of the EJC become concentrated at sites of Ad transcription. The results show that in noninfected cells, REF/Aly, Y14, SRm160, UAP56, RNPS1, and Magoh are localized throughout the nucleoplasm with higher concentration in nuclear speckles, and excluding nucleoli (Fig. 1G; data not shown), as previously described (Blencowe et al. 1998; Loyer et al. 1998; Gatfield et al. 2001; Le Hir et al. 2001a; Rodrigues et al. 2001). At 14–16 h post-infection, each of these proteins is predominantly concentrated in ring-like structures (Fig. 1H–L). Double-labeling experiments confirm that EJC proteins colocalize with spliced Ad mRNA and spliceosomal snRNPs at the ring-like structures formed by nascent viral transcripts (Fig. 2A–C). In contrast to REF/Aly, Y14, SRm160, UAP56, RNPS1, and Magoh, the protein DEK appears more homogeneously distributed throughout the nucleoplasm with less pronounced accumulation in either nuclear speckles in noninfected cells or ring-like structures in Ad-infected cells (Fig. 2D–F).

Unlike EJC components, the proteins NXF1/TAP, p15, and Dbp5 do not accumulate in nuclear speckles of noninfected cells, but rather, distribute homogeneously throughout the nucleoplasm with higher concentration at the nuclear rim (Schmitt et al. 1999; Bachi et al. 2000; Herold et al. 2000). Following treatment with TritonX-100 prior to fixation, the nucleoplasmic pool of NXF1/TAP, p15, and Dbp5 is largely solubilized and the proteins are predominantly detected at the nuclear rim (Fig. 3A–C). This localization reflects interactions with the nuclear pore complexes, as previously described (Schmitt et al. 1999; Bachi et al. 2000; Herold et al. 2000). At 14–16 h after adenoviral infection, NXF1/TAP, p15, and Dbp5 remain predominantly localized at the nuclear rim (Fig. 3D–F). Neither of these proteins is detected in the ring-like structures that concentrate spliceosomal snRNPs (Fig. 3G–I).

In summary, our results show that REF/Aly, Y14, SRm160, UAP56, RNPS1, and Magoh are all recruited to sites of adenoviral mRNA synthesis in the nucleus of in-

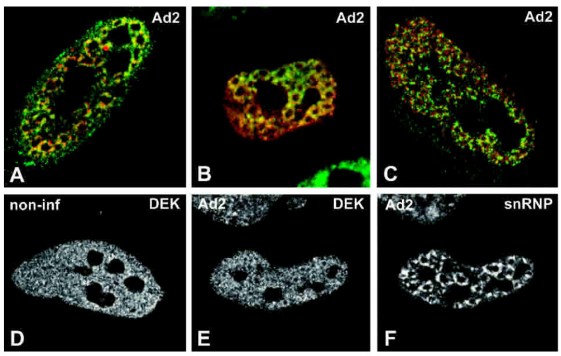


FIGURE 2. Colocalization of EJC proteins and spliceosomal snRNPs in adenovirus-infected cells. (*A*–*C*) A superimposition of red and green images corresponding to double-labeling experiments performed in HeLa cells infected with adenovirus for 14–16 h. The distribution of EJC proteins is compared with that of spliceosomal snRNPs (*A*,*B*) and spliced viral major late mRNAs (*C*). (*A*) Cells were double labeled with antibodies directed against REF/Aly (green staining) and Sm proteins (red staining). (*B*) Cells expressing zzY14 (red staining) were immunolabeled with an antibody specific for the U2 snRNP B′′ protein (green staining). (*C*) Cells were immunolabeled with anti-SRm160 antibodies (green staining) and hybridized with SJ1 probe (red staining). The distribution of DEK was analyzed by immunolabeling in noninfected (*D*) and infected (*E*) HeLa cells. The cell depicted in *E* was double labeled with antibody Y12, specific for Sm proteins (*F*).

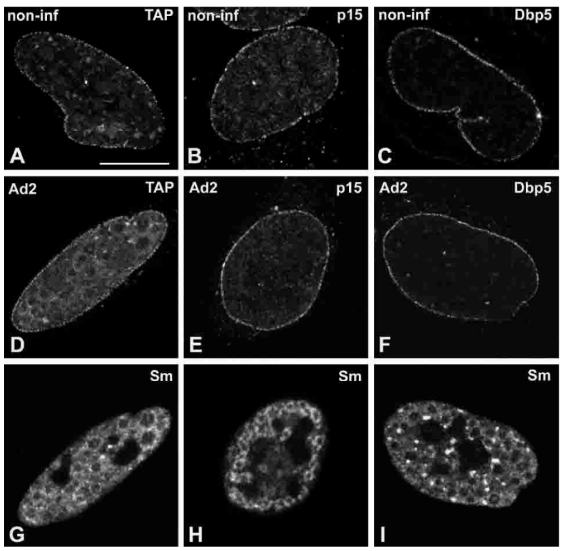


FIGURE 3. Localization of NXF1/TAP, p15, and Dbp5 in adenovirus-infected cells. HeLa cells were either noninfected (*A–C*; non-inf) or infected with adenovirus for 14–16 h (*D–I*; Ad2). Cells were transiently transfected with plasmids encoding GFP–TAP (*A*,*D*), zzp15 (*B*,*E*), and GFP–Dbp5 (*C*,*F*). The cells depicted in *D*, *E*, and *F* were double labeled with antibody Y12, specific for Sm proteins (*G*,*H*,*I*, respectively). Bar, 10 μm.

fected HeLa cells. In contrast, NXF1/TAP, p15, and Dbp5 are not detected at sites of viral transcription.

Components of the EJC, but not NXF1/TAP or p15, are recruited to sites of β-globin transcription in the nucleus of MEL cells

Having shown that EJC components accumulate at sites of adenoviral transcription, we next sought to visualize recruitment of these proteins to nascent cellular transcripts. As a model system, we used murine erythroleukemia (MEL) cells stably transfected with the human β-globin gene. The clone used in this study (MELβWT) harbors ~14 copies of the transgene as a tandem array (Custódio et al. 1999). Detection of the human β-globin transcripts by FISH was performed with a probe complementary to the transcribed sequence of the gene (RNA probe, Fig. 4A, a). Expression of the human β-globin transgenes was induced by dimethylsulfoxide, which triggers terminal erythroid differentiation of the MEL cells (Antoniou 1991). In cells induced for 2 d, nascent β-globin transcripts are detected as a focus in the nucleus (Fig. 4A, b). After induction for 4 d, the cells contain a focus in the nucleus corresponding to nascent β-globin transcripts and additional cytoplasmic staining (Fig. 4A, c). To determine whether EJC proteins colocalize with nascent β-globin transcripts, MELβWT cells were induced to differentiate and sequentially hybridized with the RNA probe and immunolabeled with antibodies against SRm160 and REF (Fig. 4B, a and b). Alternatively, cells were transfected with plasmids encoding tagged versions of Y14 and UAP56 (Fig. 4B, c and d). The results indicate that each EJC protein colocalizes with the β-globin RNA focus in the nucleus. Quantification of the fluorescence intensity along a line that spans the nucleus across the RNA focus shows that the concentration of EJC proteins at the site of β-globin transcription is higher than in surrounding areas of the nucleoplasm (Fig 4B, a′–d′). Similar results were obtained for MEL cells labeled with antibodies against Sm proteins (Fig. 4C, a and a′) or U2snRNP B′′ protein (Fig. 4C, b and b′).

Next, MELβWT cells expressing tagged NXF1/TAP or p15 were hybridized with the β-globin RNA probe (Fig. 4D, a and b). NXF1/TAP and p15 are both detected homogeneously distributed throughout the nucleoplasm with higher accumulation at the nuclear rim, reflecting interactions with the nuclear pore complexes. Quantification of the fluorescence intensity along a line across the RNA focus clearly shows that NXF1/TAP and p15 do not concentrate at the site of β-globin transcription (Fig. 4D, a′ and b′).

EJC proteins are not recruited by splicing-defective mutant β-globin transcripts

Current biochemical evidence indicates that EJC proteins are deposited onto mRNAs as a consequence of splicing (Le Hir et al. 2000a,b, 2001b; Kataoka et al. 2001; Kim et al. 2001). To study the role of splicing on recruitment of EJC proteins to sites of transcription in the nucleus, we have made use of MEL cells stably transfected with a human β-globin mutant gene that contains a normal first intron but is devoid of the second intron (termed βIVSI).

To analyze the processing of βIVSI pre-mRNAs in the nucleus of MEL cells, we performed RNase protection assays using nuclear RNA fractions from 4-day-induced cells and ^{32}P-labeled antisense RNA probes (Fig. 5A). The protection products were quantified by PhosphorImager, and the results show that for the wild type β-globin transcripts (βWT), the percentage of splicing of introns I and II is 65 and 68, respectively (Fig. 5B, lane 1). Thus, splicing of both introns has occurred in the majority of wild type β-globin transcripts (βWT) present in the nucleus. Consistent with the genomic deletion of the second intron in the βIVSI mutant, no intron II sequences are detected in the corresponding transcripts (Fig. 5B, lane 2). Unexpectedly, the first intron remains unspliced in all βIVSI transcripts (Fig. 5B, lane 2). Furthermore, and according to data reported previously (Collis et al. 1990; Antoniou et al. 1998), the efficiency of 3′-end cleavage is reduced from 93% in βWT RNA to 32% in βIVSI transcripts (Fig. 5B, lanes 3,4).

We have previously shown that βIVSI transcripts are not exported to the cytoplasm, being retained at the site of

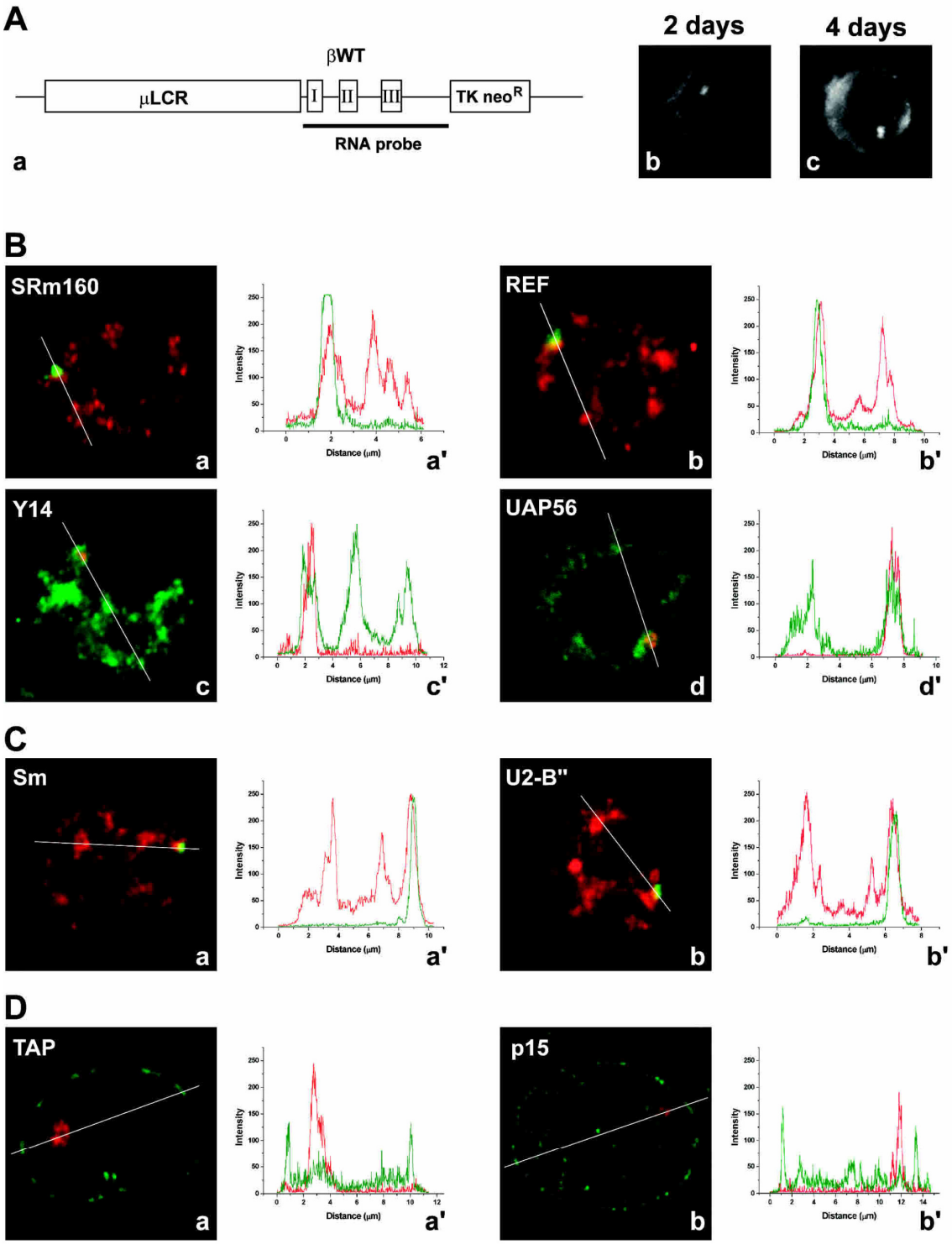


FIGURE 4. (Legend on next page)

transcription (Custódio et al. 1999). We consider it unlikely that the reduced cleavage efficiency is the primary cause of nuclear retention, as another β-globin splicing mutant possessing a 5′-splice site mutation (GT→AC) in intron II, is correctly cleaved at normal rates, yet is also stalled at the site of transcription (Custódio et al. 1999). However, we cannot rule out the possibility that inefficient 3′-end processing contributes to retention of nascent β-globin RNA at the site of transcription.

Having determined that there is no splicing of βIVSI transcripts in the nucleus of MEL cells, we next performed double-labeling experiments using a probe to detect β-globin RNA and antibodies or GFP to detect SRm160, REF, UAP56, and Sm. The results clearly show no significant colocalization of the RNA and protein signals (Fig. 6a–d). Quantification of the fluorescence intensity along a line that spans the nucleus across the RNA focus confirms that there is no accumulation of EJC or Sm proteins at the site of transcription of the mutant β-globin gene (Fig. 6a′–d′). This suggests that, in the absence of efficient pre-mRNA processing, both spliceosomal components and EJC proteins fail to be properly assembled onto nascent β-globin transcripts in the nucleus.

DISCUSSION

In this work, we used a microscopy approach to study recruitment of the EJC to nascent transcripts in the nucleus of mammalian cells. We find that EJC proteins colocalize with protein components of the spliceosome at sites of transcription and splicing in the nucleoplasm, providing in vivo evidence that the EJC binds cotranscriptionally to mRNPs. We further report that mutant β-globin pre-mRNAs that are neither spliced nor released from the site of transcription fail to recruit spliceosome snRNPs and EJC proteins. This suggests that splicing of β-globin pre-mRNAs is required for efficient recruitment of the EJC and subsequent targeting of the resulting mRNP to the nuclear export pathway.

We show that the EJC proteins SRm160, RNPS1, Y14, Magoh, REF/Aly, and UAP56 accumulate in the nucleus at sites of abundant adenoviral and β-globin pre-mRNA synthesis. This is consistent with reports that the yeast homolog of REF/Aly, Yra1, associates with transcribed chromatin (Lei et al. 2001; Lei and Silver 2002a), and that both Yra1 and Sub2 (the yeast counterpart of UAP56) associate with the THO transcription complex (Strasser et al. 2002). A different result is observed for DEK, which appears more homogeneously distributed throughout the nucleoplasm, with no clear concentration at transcription sites. Remarkably, recent immunopurification analysis of nuclear mRNPs detected all EJC components except DEK (Lykke-Andersen et al. 2001; Lejeune et al. 2002). The failure to detect DEK using both microscopic and biochemical approaches suggests that either DEK associates loosely with the EJC or it is not part of this complex in vivo.

In contrast to EJC proteins, Dbp5, NXF1/TAP, and p15 fail to concentrate in the nucleus at sites of abundant pre-mRNA synthesis. Dbp5 was identified originally in yeast as a DEAD-box helicase implicated in mRNA export (for review, see Reed and Hurt 2002). Although early studies suggested that the function of metazoan Dbp5 is conserved, more recent analysis using RNA interference indicate that this protein is not essential for mRNA export in *Drosophila* (Gatfield et al. 2001). Like the yeast protein, human Dbp5 (hDbp5) shuttles between the nucleus and the cytoplasm, but at steady state, it is detected mainly in the cytoplasm (Schmitt et al. 1999). The Dbp5/hDbp5 protein appears enriched at the nuclear periphery, where it interacts with nucleoporins located at the cytoplasmic fibrils of nuclear pore complexes (Hodge et al. 1999; Schmitt et al. 1999; Strahm et al. 1999). Here, we observe that Dbp5 distribution is not altered upon infection of HeLa cells with adenovirus (Fig. 3C,F). No Dbp5 labeling was detected at sites of nascent viral pre-mRNAs, suggesting that in mammalian cells, this protein is recruited to mRNPs shortly before or after release from the sites of transcription. However, a recent report indicates that a Dbp5 homolog in the dipteran

FIGURE 4. Localization of EJC proteins, NXF1/TAP, and p15 in the nucleus of MEL cells. (*A*, *a*) Schematic representation of the wild-type human β-globin construct. The wild-type human β-globin gene (βWT) is within the microlocus control region (μLCR) expression cassette (Collis et al. 1990). The TK *neo*^{R} gene confers resistance to G418 in stable transfected MEL cells. The probe used to detect the human β-globin transcripts by FISH (RNA probe) corresponds to a DNA fragment complementary to the transcribed sequence of the gene. (*b*,*c*) Cells induced for 2 and 4 d, and hybridized with the RNA probe. (*B*) EJC proteins were detected either by indirect immunofluorescence using specific antibodies (*a*,*b*), or by transfection with zz or GFP-tagged constructs (*c*,*d*). (*a*,*b*) MELβWT were induced to differentiate for 3 d, hybridized with the RNA probe (green), and labeled with antibodies (red) against SRm160 (*a*) or REF (*b*). (*c*,*d*) Cells were first transfected with either zzY14 (*c*) or GFP–UAP56 (*d*), induced to differentiate for 2 d, and hybridized with the RNA probe directly labeled with Cy3 (red). The graphics in *a*′, *b*′, *c*′, and *d*′ correspond to a quantitative measurement of colocalization of the EJC proteins at the sites of human β-globin transcription. A line scan was made across the nucleus, including the site of β-globin transcription as indicated in the figures. The color of the lines in the graphics matches the color of the detection of RNA and protein in the cell. (*C*) MELβWT were induced to differentiate for 3 d, hybridized with the RNA probe (green), and labeled with antibodies (red) against either U2 snRNP B″ with mAb 4G3 (*b*) or Sm with mAb Y12 (*a*). The graphics in *a*′, *b*′, correspond to a quantitative measurement of colocalization of spliceosomal proteins (red) at the sites of human β-globin transcription (green). (*D*) MELβWT were transfected with either GFP–TAP (*a*) or zzp15 (*b*), induced to differentiate for 2 d, and hybridized with the RNA probe directly labeled with Cy3 (red). The zz tag (green staining) was detected after in situ hybridization by indirect immunofluorescence with an antibody against protein A. The graphics in *a*′ and *b*′ correspond to a quantitative measurement of colocalization of NXF1/TAP or p15 (green) at the sites of human β-globin transcription (red).

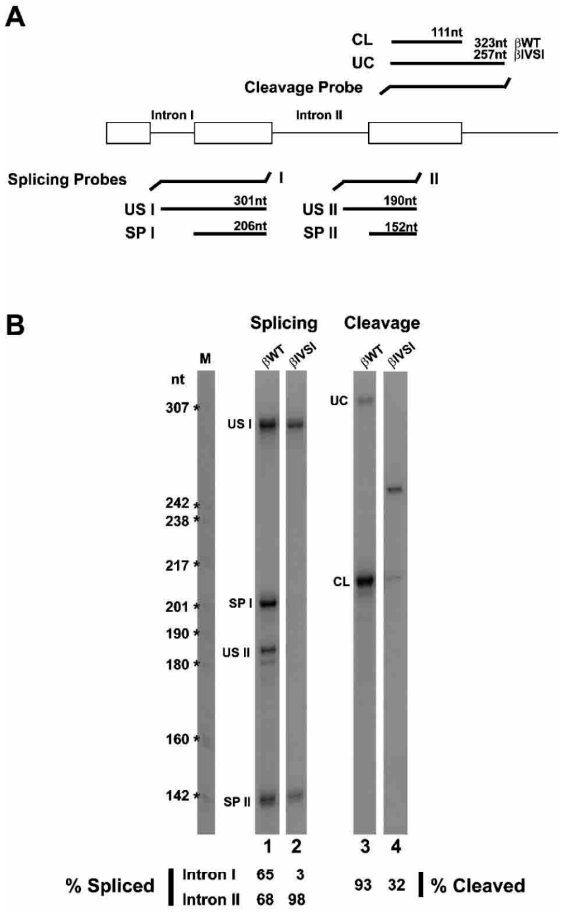


FIGURE 5. RNase protection assays. (*A*) Schematic representation of the RNase protection probes used to analyze splicing and 3′-end cleavage of the human β-globin transcripts. The predicted RNase protection products are shown for each probe. (*B*) Nuclear RNA fractions (3 μg) of βwild-type (lanes *1*,*3*) and βIVSI (lanes *2*,*4*) cells induced to differentiate for 4 d were analyzed by RNase protection using either the splicing protection probes I and II simultaneously (lanes *1*,*2*) or the cleavage probe (lanes *3*,*4*). The identity of each RNA species is indicated. The uncleaved product is shorter for βIVSI, because this transgene terminates 45 bp past the poly(A)-addition site, whereas the βWT terminates at 1800 pb. This difference in the extent of 3′ sequences does not in itself compromise the efficiency of 3′ end formation (Antoniou et al. 1998). The amounts of unspliced (US), spliced (SP), uncleaved (UC), and cleaved (CL) RNAs were quantified by PhosphorImager and the values normalized for U content. The percentage of spliced intron I and intron II is indicated *below* lanes *1* and *2*, and the percentage of cleaved RNA is indicated *below* lanes *3* and *4*.

Chironomus tentans (Ct-Dbp5) binds to the Balbiani ring pre-mRNP cotranscriptionally and accompanies the mRNP to and through the nuclear pores (Zhao et al. 2002). An involvement of yeast Dbp5 during the early steps of transcription was also suggested on the basis of genetic and physical interactions with the transcriptional factor IIH (Estruch and Cole 2003). Due to the sensitivity limits of our assay, we cannot exclude that some Dbp5 molecules may associate with nascent transcripts.

The heterodimer formed by vertebrate NXF1/TAP and NXT1/p15 represents to date the most important receptor involved in export of mRNA from the nucleus to the cytoplasm (Izaurralde 2002; Lei and Silver 2002b; Reed and Hurt 2002; Cullen 2003; Stutz and Izaurralde 2003). Several EJC proteins, including REF/Aly, Y14, and Magoh can bind to NXF1/TAP, suggesting that these factors may act as adaptors that recruit the export receptor to mRNAs (Dreyfuss et al. 2002; Reed and Hurt 2002). However, our observation that NXF1/TAP and NXT1/p15 do not accumulate at sites of transcription suggests that binding of the EJC proteins to the mRNA is not sufficient to recruit the TAP-p15 heterodimer. This is consistent with the recent finding that REF/Aly and other components of the EJC are dispensable for export of bulk mRNA in *Drosophila* (Gatfield and Izaurralde 2002) and *Caenorhabditis elegans* (Longman et al. 2003; MacMorris et al. 2003), indicating that additional adaptor proteins mediate the interaction between NXF1/TAP and mRNAs in metazoan. SR proteins were recently identified as adaptors that, in addition to REF/Aly, can mediate the interaction between NXF1/TAP and cellular mRNAs (Huang et al. 2003).

Although REF/Aly can interact directly with NXF1/TAP (Strasser and Hurt 2000; Stutz et al. 2000), and NXF1/TAP coimmunopurifies with EJC proteins in nuclear mRNP fractions (Lejeune et al. 2002), there is no evidence that the TAP-p15 heterodimer is recruited to mRNAs cotranscriptionally. In fact, NXF1/TAP was reported to bind very weakly to purified spliced mRNPs in vitro (Zhou et al. 2000). Our observation that NXF1/TAP and p15 do not concentrate at the sites of pre-mRNA transcription, therefore, favors the view that this heterodimer is recruited efficiently at a later stage in the export pathway.

At present, it is well established that EJC proteins are deposited onto mRNAs as a consequence of splicing (Le Hir et al. 2000a,b, 2001b; Kataoka et al. 2001; Kim and Dreyfus 2001). More recent studies described the timing of EJC assembly on spliced mRNA (Lejeune et al. 2002; Reichert et al. 2002). REF/Aly was shown to interact with pre-mRNA prior to spliceosome assembly, whereas Y14, Magoh, RNPS1, UAP56, and SRm160 are found in intermediate-containing spliceosomes (Reichert et al. 2002). Upon exon ligation, association of RNPS1, UAP56, and SRm160 are destabilized, whereas REF/Aly, Y14, and Magoh remain stably bound to the spliced exons (Reichert et al. 2002).

In a previous study, we have shown that transcripts encoded by a human β-globin mutant gene (βIVSI) that is devoid of the second intron fails to be released from the site of transcription (Custódio et al. 1999). Although pre-mRNAs encoded by this mutant contain a normal first intron, RNase protection assays show that splicing of this intron is almost completely inhibited in βIVSI transcripts (Fig. 5B).

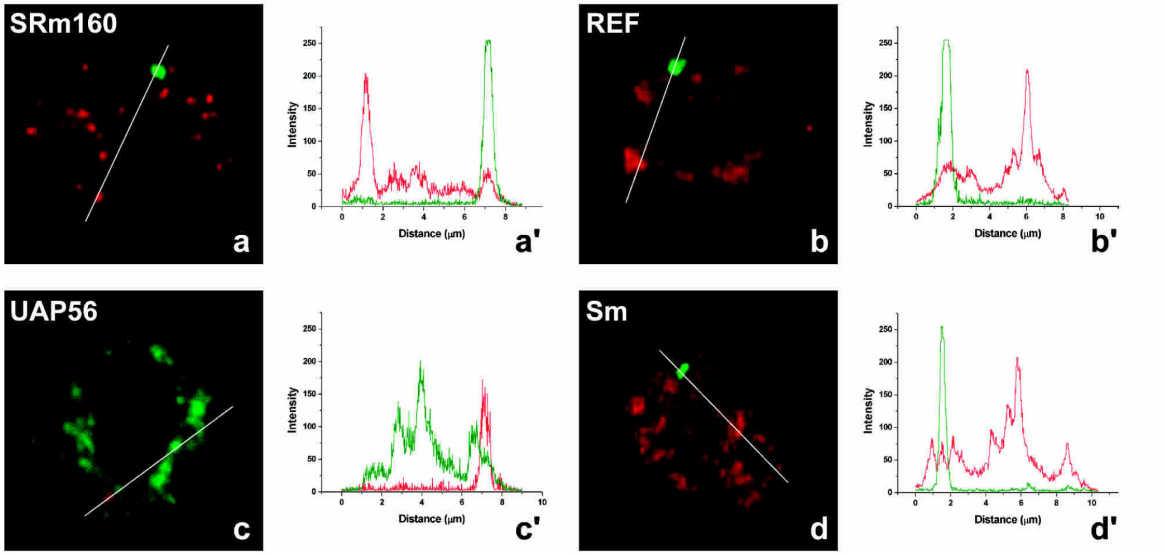


FIGURE 6. EJC proteins and spliceosomal snRNPs do not colocalize with mutant β-globin nascent transcripts. MELβIVSI were induced to differentiate for 3 d, hybridized with the RNA probe (green), and labeled with antibodies against SRm160 (*a*), REF (*b*), or Sm proteins (*d*; red). (*c*) Cells were first transfected with GFP–UAP56 (green), induced to differentiate for 2d, and hybridized with the RNA probe directly labeled with Cy3 (red). The graphics in *a′*, *b′*, *c′*, and *d′* correspond to a quantitative measurement of colocalization of the EJC proteins at the sites of mutant β-globin transcription. A line scan was made across the nucleus as indicated in the figures, and the color of the lines in the graphics matches the color of the detection of RNA and protein in the cell.

Thus, the βIVSI mutant gene generates pre-mRNAs that are not spliced. Notably, βIVSI transcripts conserve the normal sequence spanning the second and third exons, but this exon–exon junction is no longer the product of a splicing reaction. In clear contrast with the results observed in MEL cells expressing normal human β-globin transcripts, βIVSI pre-mRNAs do not colocalize with SRm160, REF, UAP56, or Sm proteins. This argues that the accumulation of EJC proteins at transcription sites requires efficient pre-mRNA processing. However, there is previous evidence indicating that REF/Aly can associate with the region 20–24 nucleotides upstream of exon–exon junctions independent of splicing (Reichert et al. 2002), that UAP56 binds to nascent transcripts independent of the localization of introns (Kiesler et al. 2002), and that both REF/Aly and UAP56 are part of the TREX (transcription/export) complex, which is thought to be recruited during transcription (Strasser et al. 2002). Possibly, in our microscopic assay we fail to visualize proteins that are loosely associated with the RNA and detect only stable complexes. According to this view, our results would argue that the initial binding of REF/Aly and UAP56 to nascent transcripts is not sufficient for stable EJC assembly.

In addition to being splicing deficient, βIVSI transcripts are not properly 3′-end cleaved. This is consistent with recent findings showing that splicing stimulates mRNA biogenesis by enhancing mRNA 3′-end processing (Lu and Cullen 2003; Nott et al. 2003). Because this effect appears to be mediated by the EJC (Wiegand et al. 2003), failure of βIVSI transcripts to recruit EJC proteins may contribute to their inefficient cleavage.

In conclusion, we find that EJC proteins colocalize with protein components of the spliceosome at sites of transcription and splicing in the nucleoplasm, providing in vivo evidence that the EJC binds cotranscriptionally to mRNPs. We further report that mutant β-globin pre-mRNAs that are neither correctly processed nor released from the site of transcription fail to concentrate spliceosome snRNPs and EJC proteins. This suggests that processing of β-globin pre-mRNAs is required for efficient recruitment of the EJC and also for subsequent targeting of the resulting mRNP to the nuclear export pathway.

MATERIALS AND METHODS

HeLa cell culture, transfections, and Adenovirus infection

HeLa cells from ECACC were cultured as monolayers in minimum essential medium (MEM, GIBCO-BRL) supplemented with 10% fetal calf serum (FCS, GIBCO-BRL) and nonessential amino acids (Invitrogen Corporation). Experiments were performed with cells grown on 10 × 10 mm glass coverslips. Transfections were performed on subconfluent cells using FuGENE 6 reagent (Roche Biochemicals), and cells were fixed for analysis 24 h after transfection. Infections with Adenovirus Type 2 (Ad2, strain wt900)

were performed as described previously (Pombo et al. 1994). Briefly, cells were incubated with Ad2 for 2 h in serum-free medium, and then supplemented with 10% FCS. For simultaneous Ad2 infection and transfection, the transfection mixture was added to the cells immediately after the addition of the FCS. For in situ detection of sites of transcription, the cells were incubated for 1 h in culture medium supplemented with 2 mM BrU (Sigma-Aldrich), and fixed for analysis.

MEL cell culture and transfections

The maintenance and induction of erythroid differentiation of the MEL cell lines were described previously (Antoniou 1991; Custódio et al. 1999). Transient transfections were performed with either FuGENE 6 reagent (Roche Biochemical) or Tfx-50 reagent (Promega). The erythroid differentiation was induced 5–6 h after transfection by the addition of 2% (v/v) DMSO to the culture medium, and allowed to proceed for 48 h.

DNA clones for transient transfections

The plasmids used for expression in HeLa and MEL cells were kindly provided by Elisa Izaurralde (EMBL-Heidelberg). NXF1/TAP, RNPS1, UAP56, Magoh, and Dbp5 full-length cDNAs were cloned in pEGFP (Clontech). Y14 and p15 cDNAs were cloned in the pRN3zz vector. The zz tag was detected by indirect immunofluorescence with an antibody against protein A.

In situ hybridization

HeLa cells were grown on glass coverslips, and MEL cells were washed in serum-free medium and allowed to adhere onto poly-L-lysine (Sigma) coated glass coverslips. Cells were fixed and permeabilized according to one of the following alternative protocols: (1) permeabilized with 0.5% Triton X-100 in CSK buffer (Fey et al. 1986) containing 0.1 mM PMSF for 1 min on ice, and fixed with 3.7% paraformaldehyde in CSK buffer for 10 min at room temperature; (2) fixed in 3.7% paraformaldehyde in PBS for 10 min, and permeabilized with 0.5% Triton X-100 in PBS for 10 min at room temperature.

Spliced major late adenoviral mRNAs were visualized using a 5′-biotinylated DNA oligonucleotide probe, as previously described (Bridge et al. 1996). Oligonucleotides are particularly suited to identify spliced mRNA molecules, as probes complementary to exon–exon splice junctions form unstable hybrids with the unspliced primary transcripts. The probe (termed SJ1) has the following sequence: CAACCGCGAGCCCAACAGCTG. All in situ hybridization procedures were based on the protocols previously described (Bridge et al. 1996), and the hybrids were detected with Cy3 avidin (1/200, Jackson ImmunoReseach Labs, Inc).

To detect the human β-globin transcripts, MEL cells were hybridized as previously described (Custódio et al. 1999). The probe used consists of a plasmid containing the genomic sequence of the human β-globin gene labeled with either digoxigenin-11-dUTP (Roche) or Cy3-AP3-dUTP (Amersham Pharmacia) by nick translation. Digoxigenin detection was either with Fluorescein-conjugated sheep anti-digoxigenin (1/100, Roche), followed by Alexa-Fluor488-conjugated goat anti-Fluorescein (1/200, Molecular Probes), or Cy3 conjugated mouse anti-digoxin (1/250, Jackson ImmunoReseach Labs, Inc.). After the detection steps, cells were fixed with 1% formaldehyde in PBS for 10 min, washed with PBS, and processed for indirect immunofluorescence.

Immunofluorescence

The following primary antibodies were used for immunofluorescence. Rabbit polyclonals directed against (1) SRm160 (1/250, Blencowe et al. 1998); (2) DEK (1/300, Fornerod et al. 1995); (3) REF (1/100, Rodrigues et al. 2001); (4) protein A (1/2000, Sigma); (5) adenoviral protein DBP (Linné et al. 1977). Mouse monoclonals directed against (1) Sm antigen of SnRNPs (mAb Y12; Lerner et al. 1981); (2) U2 snRNP specific protein B′′ (mAb 4G3; Habets et al. 1989). Bromo-uridine incorporation was detected with a sheep polyclonal antibody directed against BrdU (1/200, Abcam).

The secondary antibodies used were AlexaFluor488-conjugated goat anti-rabbit IgG (1/200), TRITC-conjugated donkey anti-rabbit IgG (1/100, Jackson ImmunoResearch Labs, Inc.), Alexa-Fluo488-conjugated goat anti-mouse IgG (1/200, Jackson), Cy3-conjugated goat anti-mouse IgG (1/300, Jackson), FITC-conjugated goat anti-human IgG (1/100, Jackson), Cy5-conjugated rabbit anti-human IgG (1/100, Jackson), and FITC-conjugated donkey anti-sheep IgG (1/100, Jackson).

Microscopy

Images were acquired on a Zeiss LSM 510 confocal microscope using the PlanApochromat 63×/1.4 objective. FITC and Alexa-Fluor488 fluorescence was detected using the 488-nm line of the argon ion laser. The 543-nm line of the helium-neon laser was used to excite Cy3 and TRITC and the 633-nm line to excite Cy5.

RNase protection assay

Nuclear and cytoplasmic fractionation of MEL cells was as previously described (Antoniou et al. 1998). The nuclei were lysed with 4 M guanidinium thiocyanate, 25 mM tri-sodium citrate, 0.5% (w/v) *N*-lauryl sarcosine, 0.1 mM dithiothreitol (Huang and Carmichael 1996), and the homogenate sonicated (30 pulses using a small-diameter probe) to shear the DNA. The RNA was purified by phenol:chloroform (2.8:1) extraction, isopropanol and ethanol precipitations, and DNase digestion (RNase free, Roche). Human β-globin RNase protection probes (McCracken et al. 1997) were prepared by in vitro transcription with T7 RNA polymerase in the presence of [α-^{32}P]UTP, and gel-purified prior to use. RNase protection assays were performed as previously described (McCracken et al. 2002). Briefly, 3 μg of nuclear RNA were incubated with the antisense RNA probe overnight at 50°C, and the hybridization products were digested with a mixture of RNase T1 (5 μg/mL) and RNase A (0.5 μg/mL) at 37°C for 1 h. The protected fragments were resolved on a 6% denaturing polyacrylamide gel, and the intensity of the bands quantified using a BioRad PhosphorImager. Following quantification of each gel band, background was subtracted and the values normalized for U content of the protected probe fragment.

ACKNOWLEDGMENTS

We thank José Braga, José Rino, Helena Pina, and Maria do Carmo Silva for technical support and Susan McCracken (University of Toronto) for help with the RNase protection assay. We thank Elisa Izaurralde (EMBL-Heidelberg) for providing plasmids and antibodies. This study was supported by a grant (POCTI/MGI/36547/00) from Fundação para a Ciência e Tecnologia (FCT), Portugal, and by the European Commission (QLG2-CT-2001-01554).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Received December 10, 2003; accepted January 12, 2004.

REFERENCES

Alexiadis, V., Waldmannm, T., Andersen, J., Mann, M., Knippers, R., and Gruss, C. 2000. The protein encoded by the proto-oncogene DEK changes the topology of chromatin and reduces the efficiency of DNA replication in a chromatin-specific manner. *Genes & Dev.* **14:** 1308–1312.

Antoniou, A. 1991. Induction of erythroid-specific expression in murine erythroleukemia (MEL) cells. In *Methods in molecular biology: Gene transfer and expression protocols.* (ed. E.J. Murray), pp. 421–434. The Human Press, Inc., Clinfton, N.J.

Antoniou, M., Geraghty, F., Hurst, J., and Grosveld, F. 1998. Efficient 3′-end formation of human β-globin mRNA in vivo requires sequences within the last intron but occurs independently of the splicing reaction. *Nucleic Acids Res.* **26:** 721–729.

Bachi, A., Braun, I.C., Rodrigues, J.P., Pante, N., Ribbeck, K., von Kobbe, C., Kutay, U., Wilm, M., Gorlich, D., Carmo-Fonseca, M., et al. 2000. The C-terminal domain of TAP interacts with the nuclear pore complex and promotes export of specific CTE-bearing RNA substrates. *RNA* **6:** 136–158.

Bauren, G. and Wieslander, L. 1994. Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* **76:** 183–192.

Bauren, G., Jiang, W.Q., Bernholm, K., Gu, F., and Wieslander, L. 1996. Demonstration of a dynamic, transcription-dependent organization of pre-mRNA splicing factors in polytene nuclei. *J. Cell. Biol.* **133:** 929–941.

Beyer, A.L. and Osheim, Y.N. 1988. Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes & Dev.* **2:** 754–765.

Blencowe, B.J., Issner, R., Nickerson, J.A., and Sharp, P.A. 1998. A coactivator of pre-mRNA splicing. *Genes & Dev.* **12:** 996–1009.

Bridge, E., Xia, D.X., Carmo-Fonseca, M., Cardinali, B., Lamond, A.I., and Pettersson, U. 1995. Dynamic organization of splicing factors in adenovirus-infected cells. *J. Virol.* **69:** 281–290.

Bridge, E., Riedel, K.U., Johansson, B.M., and Pettersson, U. 1996. Spliced exons of adenovirus late RNAs colocalize with snRNP in a specific nuclear domain. *J. Cell. Biol.* **135:** 303–314.

Collis, P., Antoniou, M., and Grosveld, F. 1990. Definition of the minimal requirements within the human β-globin gene and the dominant control region for high level expression. *EMBO J.* **9:** 233–240.

Cullen, B.R. 2003. Nuclear RNA export. *J. Cell. Sci.* **116:** 587–597.

Custódio, N., Carmo-Fonseca, M., Geraghty, F., Pereira, H.S., Grosveld, F., and Antoniou, M. 1999. Inefficient processing impairs release of RNA from the site of transcription. *EMBO J.* **18:** 2855–2866.

Dreyfuss, G., Kim, V.N., and Kataoka, N. 2002. Messenger-RNA-binding proteins and the messages they carry. *Nat. Rev. Mol. Cell. Biol.* **3:** 195–205.

Estruch, F. and Cole, C.N. 2003. An early function during transcription for the yeast mRNA export factor Dbp5p/Rat8p suggested by its genetic and physical interactions with transcription factor IIH components. *Mol. Biol. Cell* **14:** 1664–1676.

Faulkner, N.E., Hilfinger, J.M., and Markovitz, D.M. 2001. Protein phosphatase 2A activates the HIV-2 promoter through enhancer elements that include the pets site. *J. Biol. Chem.* **276:** 25804–25812.

Fey, E.G., Krochmalnic, G., and Penman, S. 1986. The nonchromatin substructures of the nucleus: The ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy. *J. Cell. Biol.* **102:** 1654–1665.

Fleckner, J., Zhang, M., Valcarcel, J., and Green, M.R. 1997. U2AF65 recruits a novel human DEAD box protein required for the U2 snRNP-branchpoint interaction. *Genes & Dev.* **11:** 1864–1872.

Flint, S.J. 1986. Regulation of adenovirus mRNA formation. *Adv. Virus Res.* **31:** 169–228.

Fornerod, M., Boer, J., van Baal, S., Jaegle, M., von Lindern, M., Murti, K.G., Davis, D., Bonten, J., Buijs, A., and Grosveld, G. 1995. Relocation of the carboxy terminal part of CAN from the nuclear envelope to the nucleus as a result of leukemia-specific chromosome rearrangements. *Oncogene* **10:** 1739–1748.

Fribourg, S., Gatfield, D., Izaurralde, E., and Conti, E. 2003. A novel mode of RBD-protein recognition in the Y14-Mago complex. *Nat. Struct. Biol.* **10:** 433–439.

Gama-Carvalho, M., Krauss, R.D., Chiang, L., Valcarcel, J., Green, M.R., and Carmo-Fonseca, M. 1997. Targeting of U2AF65 to sites of active splicing in the nucleus. *J. Cell. Biol.* **137:** 975–987.

Gama-Carvalho, M., Condado, I., and Carmo-Fonseca, M. 2003. Regulation of adenovirus alternative RNA splicing correlates with a re-organization of splicing factors in the nucleus. *Exp. Cell. Res.* **289:** 77–85.

Gatfield, D. and Izaurralde, E. 2002. REF1/Aly and the additional exon junction complex proteins are dispensable for nuclear mRNA export. *J. Cell. Biol.* **159:** 579–588.

Gatfield, D., Le Hir, H., Schmitt, C., Braun, I.C., Kocher, T., Wilm, M., and Izaurralde, E. 2001. The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in *Drosophila*. *Curr. Biol.* **11:** 1716–1721.

Greber, U.F., Willetts, M., Webster, P., and Helenius, A. 1993. Stepwise dismantling of adenovirus 2 during entry into cells. *Cell* **75:** 477–486.

Habets, W.J., Hoet, M.H., De Jong, B.A., Van der Kemp, A., and Van Venrooij, W.J. 1989. Mapping of B cell epitopes on small nuclear ribonucleoproteins that react with human autoantibodies as well as with experimentally-induced mouse monoclonal antibodies. *J. Immunol.* **143:** 2560–2566.

Hachet, O. and Ephrussi, A. 2001. *Drosophila* Y14 shuttles to the posterior of the oocyte and is required for oskar mRNA transport. *Curr. Biol.* **11:** 1666–1674.

Herold, A., Suyama, M., Rodrigues, J.P., Braun, I.C., Kutay, U., Carmo-Fonseca, M., Bork, P., and Izaurralde, E. 2000. TAP (NXF1) belongs to a multigene family of putative RNA export factors with a conserved modular architecture. *Mol. Cell. Biol.* **20:** 8996–9008.

Hodge, C.A., Colot, H.V., Stafford, P., and Cole, C.N. 1999. Rat8p/Dbp5p is a shuttling transport factor that interacts with Rat7p/Nup159p and Gle1p and suppresses the mRNA export defect of xpo1–1 cells. *EMBO J.* **18:** 5778–5788.

Huang, S. and Spector, D.L. 1996. Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription. *J. Cell. Biol.* **133:** 719–732.

Huang, Y. and Carmichael, G.C. 1996. Role of polyadenylation in nucleocytoplasmic transport of mRNA. *Mol. Cell. Biol.* **16:** 1534–1542.

Huang, Y., Gattoni, R., Stevenin, J., and Steitz, J.A. 2003. SR splicing factors serve as adapter proteins for TAP-dependent mRNA ex-

port. *Mol. Cell* **11:** 837–843.
Izaurralde, E. 2002. A novel family of nuclear transport receptors mediates the export of messenger RNA to the cytoplasm. *Eur. J. Cell. Biol.* **81:** 577–584.
Jimenez-Garcia, L.F. and Spector, D.L. 1993. In vivo evidence that transcription and splicing are coordinated by a recruiting mechanism. *Cell* **73:** 47–59.
Kataoka, N., Yong, J., Kim, V.N., Velazquez, F., Perkinson, R.A., Wang, F., and Dreyfuss, G. 2000. Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol. Cell* **6:** 673–682.
Kataoka, N., Diem, M.D., Kim, V.N., Yong, J., and Dreyfuss, G. 2001. Magoh, a human homolog of *Drosophila* mago nashi protein, is a component of the splicing-dependent exon-exon junction complex. *EMBO J.* **20:** 6424–6433.
Kiesler, E., Miralles, F., and Visa, N. 2002. HEL/UAP56 binds cotranscriptionally to the Balbiani ring pre-mRNA in an intron-independent manner and accompanies the BR mRNP to the nuclear pore. *Curr. Biol.* **12:** 859–862.
Kim, V.N. and Dreyfus, G. 2001. Nuclear mRNA binding proteins couple pre-mRNA splicing and post-splicing events. *Mol. Cells* **12:** 1–10.
Kim, V.N., Kataoka, N., and Dreyfuss, G. 2001. Role of the nonsense-mediated decay factor hUpf3 in the splicing- dependent exon-exon junction complex. *Science* **293:** 1832–1836.
Kistler, A.L. and Guthrie, C. 2001. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. *Genes & Dev.* **15:** 42–49.
Lau, C.K., Diem, M.D., Dreyfuss, G., and Van Duyne, G.D. 2003. Structure of the y14-magoh core of the exon junction complex. *Curr. Biol.* **13:** 933–941.
Le Hir, H., Izaurralde, E., Maquat, L.E., and Moore, M.J. 2000a. The spliceosome deposits multiple proteins 20–24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J.* **19:** 6860–6869.
Le Hir, H., Moore, M.J., and Maquat, L.E. 2000b. Pre-mRNA splicing alters mRNP composition: Evidence for stable association of proteins at exon-exon junctions. *Genes & Dev.* **14:** 1098–1108.
Le Hir, H., Gatfield, D., Braun, I.C., Forler, D., and Izaurralde, E. 2001a. The protein Mago provides a link between splicing and mRNA localization. *EMBO Rep.* **2:** 1119–1124.
Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. 2001b. The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J.* **20:** 4987–4997.
Lei, E.P. and Silver, P.A. 2002a. Intron status and 3′-end formation control cotranscriptional export of mRNA. *Genes & Dev.* **16:** 2761–2766.
———. 2002b. Protein and RNA export from the nucleus. *Dev. Cell* **2:** 261–272.
Lei, E.P., Krebber, H., and Silver, P.A. 2001. Messenger RNAs are recruited for nuclear export during transcription. *Genes & Dev.* **15:** 1771–1782.
Lejeune, F., Ishigaki, Y., Li, X., and Maquat, L.E. 2002. The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: Dynamics of mRNP remodeling. *EMBO J.* **21:** 3536–3545.
LeMaire, M.F. and Thummel, C.S. 1990. Splicing precedes polyadenylation during *Drosophila* E74A transcription. *Mol. Cell. Biol.* **10:** 6059–6063.
Lerner, E.A., Lerner, M.R., Janeway Jr., C.A., and Steitz, J.A. 1981. Monoclonal antibodies to nucleic acid-containing cellular constituents: probes for molecular biology and autoimmune disease. *Proc. Natl. Acad. Sci.* **78:** 2737–2741.
Libri, D., Graziani, N., Saguez, C., and Boulay, J. 2001. Multiple roles for the yeast SUB2/yUAP56 gene in splicing. *Genes & Dev.* **15:** 36–41.
Linné, T., Jornvall, H., and Philipson, L. 1977. Purification and characterization of the phosphorylated DNA-binding protein from adenovirus-type-2-infected cells. *Eur. J. Biochem.* **76:** 481–490.
Longman, D., Johnstone, I.L., and Caceres, J.F. 2003. The Ref/Aly proteins are dispensable for mRNA export and development in *Caenorhabditis elegans*. *RNA* **9:** 881–891.
Loyer, P., Trembley, J.H., Lahti, J.M., and Kidd, V.J. 1998. The RNP protein, RNPS1, associates with specific isoforms of the p34cdc2-related PITSLRE protein kinase in vivo. *J. Cell. Sci.* **111:** 1495–1506.
Lu, S. and Cullen, B.R. 2003. Analysis of the stimulatory effect of splicing on mRNA production and utilization in mammalian cells. *RNA* **9:** 618–630.
Luo, M.L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M., and Reed, R. 2001. Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature* **413:** 644–647.
Lykke-Andersen, J., Shu, M.D., and Steitz, J.A. 2001. Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science* **293:** 1836–1839.
MacMorris, M., Brocker, C., and Blumenthal, T. 2003. UAP56 levels affect viability and mRNA export in *Caenorhabditis elegans*. *RNA* **9:** 847–857.
Mayeda, A., Badolato, J., Kobayashi, R., Zhang, M.Q., Gardiner, E.M., and Krainer, A.R. 1999. Purification and characterization of human RNPS1: A general activator of pre-mRNA splicing. *EMBO J.* **18:** 4560–4570.
McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D.L. 1997. 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes & Dev.* **11:** 3306–3318.
McCracken, S., Lambermon, M., and Blencowe, B.J. 2002. SRm160 splicing coactivator promotes transcript 3′-end cleavage. *Mol. Cell. Biol.* **22:** 148–160.
McGarvey, T., Rosonina, E., McCracken, S., Li, Q., Arnaout, R., Mientjes, E., Nickerson, J.A., Awrey, D., Greenblatt, J., Grosveld, G., et al. 2000. The acute myeloid leukemia-associated protein, DEK, forms a splicing-dependent interaction with exon-product complexes. *J. Cell. Biol.* **150:** 309–320.
Mohr, S.E., Dillon, S.T., and Boswell, R.E. 2001. The RNA-binding protein Tsunagi interacts with Mago Nashi to establish polarity and localize oskar mRNA during *Drosophila* oogenesis. *Genes & Dev.* **15:** 2886–2899.
Neugebauer, K.M. and Roth, M.B. 1997. Transcription units as RNA processing units. *Genes & Dev.* **11:** 3279–3285.
Nott, A., Meislin, S.H., and Moore, M.J. 2003. A quantitative analysis of intron effects on mammalian gene expression. *RNA* **9:** 607–617.
Philipson, L., Pettersson, U., and Lindberg, U. 1975. Molecular biology of adenoviruses. *Virol. Monogr.* **14:** 1–115.
Pombo, A. and Carmo-Fonseca, M. 1995. Interaction of adenovirus with the nucleus of the host cell. *Rev. Med. Virol.* **5:** 213–218.
Pombo, A., Ferreira, J., Bridge, E., and Carmo-Fonseca, M. 1994. Adenovirus replication and transcription sites are spatially separated in the nucleus of infected cells. *EMBO J.* **13:** 5075–5085.
Reed, R. and Hurt, E. 2002. A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell* **108:** 523–531.
Reichert, V.L., Le Hir, H., Jurica, M.S., and Moore, M.J. 2002. 5′ exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly. *Genes & Dev.* **16:** 2778–2791.
Rodrigues, J.P., Rode, M., Gatfield, D., Blencowe, B.J., Carmo-Fonseca, M., and Izaurralde, E. 2001. REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus. *Proc. Natl. Acad. Sci.* **98:** 1030–1035.
Schmitt, C., von Kobbe, C., Bachi, A., Pante, N., Rodrigues, J.P., Boscheron, C., Rigaut, G., Wilm, M., Seraphin, B., Carmo-Fonseca, M., et al. 1999. Dbp5, a DEAD-box protein required for mRNA export, is recruited to the cytoplasmic fibrils of nuclear pore complex via a conserved interaction with CAN/Nup159p. *EMBO J.* **18:** 4332–4347.
Snay-Hodge, C.A., Colot, H.V., Goldstein, A.L., and Cole, C.N. 1998. Dbp5p/Rat8p is a yeast nuclear pore-associated DEAD-box protein

essential for RNA export. *EMBO J.* **17:** 2663–2676.

Strahm, Y., Fahrenkrog, B., Zenklusen, D., Rychner, E., Kantor, J., Rosbach, M., and Stutz, F. 1999. The RNA export factor Gle1p is located on the cytoplasmic fibrils of the NPC and physically interacts with the FG-nucleoporin Rip1p, the DEAD-box protein Rat8p/Dbp5p and a new protein Ymr 255p. *EMBO J.* **18:** 5761–5777.

Strasser, K. and Hurt, E. 2000. Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J.* **19:** 410–420.

Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R., et al. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature* **417:** 304–308.

Stutz, F. and Izaurralde, E. 2003. The interplay of nuclear mRNP assembly, mRNA surveillance and export. *Trends Cell. Biol.* **13:** 319–327.

Stutz, F., Bachi, A., Doerks, T., Braun, I.C., Seraphin, B., Wilm, M., Bork, P., and Izaurralde, E. 2000. REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* **6:** 638–650.

Tennyson, C.N., Klamut, H.J., and Worton, R.G. 1995. The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. *Nat. Genet.* **9:** 184–190.

Tseng, S.S., Weaver, P.L., Liu, Y., Hitomi, M., Tartakoff, A.M., and Chang, T.H. 1998. Dbp5p, a cytosolic RNA helicase, is required for poly(A)+ RNA export. *EMBO J.* **17:** 2651–2662.

Virbasius, C.M., Wagner, S., and Green, M.R. 1999. A human nuclear-localized chaperone that regulates dimerization, DNA binding, and transcriptional activity of bZIP proteins. *Mol. Cell* **4:** 219–228.

Wiegand, H.L., Lu, S., and Cullen, B.R. 2003. Exon junction complexes mediate the enhancing effect of splicing on mRNA expression. *Proc. Natl. Acad. Sci.* **100:** 11327–11332.

Wu, Z.A., Murphy, C., Callan, H.G., and Gall, J.G. 1991. Small nuclear ribonucleoproteins and heterogeneous nuclear ribonucleoproteins in the amphibian germinal vesicle: Loops, spheres, and snurposomes. *J. Cell. Biol.* **113:** 465–483.

Xing, Y., Johnson, C.V., Dobner, P.R., and Lawrence, J.B. 1993. Higher level organization of individual gene transcription and RNA splicing. *Science* **259:** 1326–1330.

Zhang, G., Taneja, K.L., Singer, R.H., and Green, M.R. 1994. Localization of pre-mRNA splicing in mammalian nuclei. *Nature* **372:** 809–812.

Zhang, M. and Green, M.R. 2001. Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. *Genes & Dev.* **15:** 30–35.

Zhao, J., Jin, S.B., Bjorkroth, B., Wieslander, L., and Daneholt, B. 2002. The mRNA export factor Dbp5 is associated with Balbiani ring mRNP from gene to cytoplasm. *EMBO J.* **21:** 1177–1187.

Zhou, Z., Luo, M.J., Straesser, K., Katahira, J., Hurt, E., and Reed, R. 2000. The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407:** 401–405.

3. Splicing- and cleavage-independent requirement of RNA polymerase II CTD for mRNA release from the transcription site

One hypothesis to explain the retention of the pre-mRNA processing mutants at the transcription site was that the transcripts could be retained via the CTD of RNA Pol II. The big clue for this hypothesis came from reports showing that both splicing and 3' end processing factors could associate with the CTD of RNA Pol II. Since the processing mutants analysed were still able to assemble some processing machinery, the release of the transcripts could be blocked by the stalled or abnormal processing machinery associated with the CTD. In order to test this hypothesis we generated cell lines that express either the wild-type *HBB* gene or a mutant version defective in splicing and an α-amanitin resistant form of RNA Pol II LS with either 52 (wt), 31 (Δ31) or 5 (Δ5) repeats of the CTD. Surprisingly, the principal finding from this study was that wild-type *HBB* transcripts made by RNA Pol II with only 31 CTD repearts were retained at the site of transcription. We showed that the retained transcripts were correctly processed providing evidence that mRNA release from the transcription site requires the heptad repeat structure of the CTD. Based on this results we proposed that the missing heptads in the truncated CTD mutant were required for binding of proteins implicated in a final co-transcriptional maturation of spliced and 3' end cleaved mRNAs into export-competent ribonucleoprotein particles.

Work published in: J.Cell Biol. Vol.179 No.2 pp.199-207, 2007

I would like to stress that some of the results presented and discussed in this section are the product of collaborative work. Dr. Michael Antoniou gave a valuable contribution in the generation of the murine erythroleukemia (MEL) clones stably transfected with the α-amanitin resistant form of the RNA polymerase II largest subunit and Dr. Maria Vivo participated in the FISH analysis of the clones.

Splicing- and cleavage-independent requirement of RNA polymerase II CTD for mRNA release from the transcription site

Noélia Custódio,[1] Maria Vivo,[1] Michael Antoniou,[2] and Maria Carmo-Fonseca[1]

[1]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal
[2]Nuclear Biology Group, Department of Medical and Molecular Genetics, King's College London School of Medicine, Guy's Campus, Guy's Hospital, London SE1 9RT, England, UK

Eukaryotic cells have a surveillance mechanism that identifies aberrantly processed pre-mRNAs and prevents their flow to the cytoplasm by tethering them near the site of transcription. Here we provide evidence that mRNA release from the transcription site requires the heptad repeat structure of the C-terminal domain (CTD) of RNA polymerase II. The mammalian CTD, which is essential for normal co-transcriptional maturation of mRNA precursors, comprises 52 heptad repeats. We show that a truncated CTD containing 31 repeats (heptads 1–23, 36–38, and 48–52) is sufficient to support transcription, splicing, cleavage, and polyadenylation. Yet, the resulting mRNAs are mostly retained in the vicinity of the gene after transcriptional shutoff. The retained mRNAs maintain the ability to recruit components of the exon junction complex and the nuclear exosome subunit Rrp6p, suggesting that binding of these proteins is not sufficient for RNA release. We propose that the missing heptads in the truncated CTD mutant are required for binding of proteins implicated in a final co-transcriptional maturation of spliced and 3′ end cleaved and polyadenylated mRNAs into export-competent ribonucleoprotein particles.

Introduction

In eukaryotic cells, messenger precursor molecules must undergo a series of maturation events that include 5′ capping, splicing, 3′ end cleavage, and polyadenylation. During processing, nascent mRNA assembles together with RNA binding proteins into ribonucleoprotein particles (mRNPs; Aguilera, 2005; Moore, 2005). Mature particles are exported to the cytoplasm and several lines of evidence indicate that mRNPs move from the sites of transcription to the nuclear pores by random Brownian motion. As diffusion cannot be regulated, traffic control of newly synthesized mRNA molecules is thought to rely on retention at dedicated sites within the nucleus (Gorski et al., 2006). According to the current view, any failure compromising the integrity of an mRNA may cause its retention in the nucleus and trigger its degradation. There is evidence suggesting that such a surveillance mechanism operates in close proximity to the gene template (Jensen et al., 2003) and, at least in yeast, at the nuclear pore (Galy et al., 2004).

A key connection between transcription and mRNP biogenesis is provided by the C-terminal domain (CTD) of the largest subunit of RNA polymerase II (RNA Pol II LS), which binds several proteins essential for pre-mRNA processing (Bentley, 2005). The CTD of RNA Pol II LS is highly conserved, increasing in length and diversifying in structure with the complexity of organisms (Stiller and Hall, 2002). Contrasting with yeast, which contains 26 repeats of a conserved heptapeptide with the consensus sequence YSPTSPS, the mammalian CTD has 52 repeats, of which 21 obey the conserved consensus while the remainder display a variety of substitutions. Most of these nonconsensus repeats are located in the C-terminal part of the CTD (heptads 27–52; Fig. 1 A), and the last repeat (CTD52) is essential for cell viability and Pol II stability (Chapman et al., 2004). At the very C terminus, the mammalian CTD further comprises a specific 10-amino acid motif. CTD deletion analysis has shown that heptad repeats 1–15 or 1–25 support capping but not splicing or 3′ end formation, whereas heptads 27–52 plus the C-terminal 10 residues can support efficient capping, splicing, and 3′ end formation (Fong and Bentley, 2001). More recent studies have demonstrated that scrambling the 10 residues that lie C-terminal of heptad 52 impairs efficient release of RNA from the site of transcription (Bird et al., 2005). However, this mutation also reduces

Correspondence to Maria Carmo-Fonseca: carmo.fonseca@fm.ul.pt

M. Vivo's present address is Dipartimento di Biologia Strutturale e Funzionale, Università di Napoli "Federico II", Napoli 80126, Italy.

Abbreviations used in this paper: βLCR, *HBB* micro-locus control region; CMV, cytomegalovirus; CTD, C-terminal domain; EJC, exon junction complex; HA, haemagglutinin; *HBB*, human β-globin gene; *Hbb1*, murine β^{major}-globin gene; MEL, murine erythroleukemia; RNA Pol II LS, largest subunit of RNA polymerase II.

The Journal of Cell Biology, Vol. 179, No. 2, October 22, 2007 199–207
http://www.jcb.org/cgi/doi/10.1083/jcb.200612109

Supplemental Material can be found at:
http://www.jcb.org/cgi/content/full/jcb.200612109/DC1

splicing and 3′ end cleavage (Fong et al., 2003), arguing that the CTD requirement for RNA release may be a consequence of its role in promoting pre-mRNA processing.

To further investigate the role of the CTD in transcript release, we generated murine erythroleukemia (MEL) cell lines that express α-amanitin–resistant RNA Pol II LS with either full-length or truncated forms of the CTD. Our results reveal that deleting 21 C-terminal heptads of the CTD causes transcript retention at the site of transcription but without inhibiting splicing or 3′ end formation. This implies a previously unsuspected involvement of the CTD in mRNP maturation events that occur after splicing, cleavage, and polyadenylation have taken place.

Results and discussion

Deletion of the CTD to 5 heptads abolishes LCR-dependent transcriptional activation of the *Hbb1* gene

MEL cells were stably transfected with an α-amanitin–resistant form of the RNA polymerase II largest subunit (RNA Pol II LS; Bartolomei and Corden, 1987) containing either wild-type or deletion mutants of the CTD with 31 (Δ31) or 5 heptad repeats (Δ5) (Fig. 1 A; Bartolomei et al., 1988; Gerber et al., 1995). Each of these plasmids was cotransfected with a second plasmid containing the human β-globin gene (*HBB*) micro-locus control region (βLCR) and a puromycin resistance gene (Collis et al., 1990; Millevoi et al., 2002). Given the tendency of multiple copies of plasmid transgenes to co-integrate as tandem arrays, we reasoned that this cotransfection procedure should, in many cases, result in the positioning of the βLCR-bearing plasmid upstream of the α-amanitin–resistant RNA Pol II LS constructs. As the βLCR is able to activate heterologous, nonerythroid promoters (Blom van Assendelft et al., 1989; Collis et al., 1990) with the minimum requirements being a CAAT and CACCA or GC-rich (e.g., Sp1) elements (Antoniou and Grosveld, 1990; and unpublished data), this configuration should confer erythroid-specific induced transcription on the human cytomegalovirus (CMV) promoter linked to the α-amanitin–resistant RNA Pol II LS cassettes. Stably transfected clones were selected with puromycin and screened by an S1-nuclease protection assay for expression of the transfected RNA Pol II LS gene. We selected clones that showed low levels of exogenous, transgene-derived α-amanitin–resistant RNA Pol II LS expression in preinduced cells and high levels after 4 d of differentiation. Expression of the exogenous RNA Pol II LS was confirmed by Western blotting analysis with an antibody that recognizes the haemagglutinin (HA) epitope (Fig. 1 B), as previously described (Custódio et al., 2006).

We next confirmed that the exogenous RNA Pol II LS was functional. The endogenous RNA Pol II LS is degraded upon binding of α-amanitin (Nguyen et al., 1996), but the exogenous transgene-derived protein is resistant due to a single amino acid substitution that decreases its affinity for the toxin (Bartolomei and Corden, 1987). We therefore determined the ability of the different clones to transcribe endogenous murine β^{major}-globin gene (*Hbb1*) in the presence of α-amanitin. After 17 h of α-amanitin treatment, S1-nuclease protection assays revealed no signal for *Hbb1* pre-mRNA in untransfected MEL C88 cells, confirming

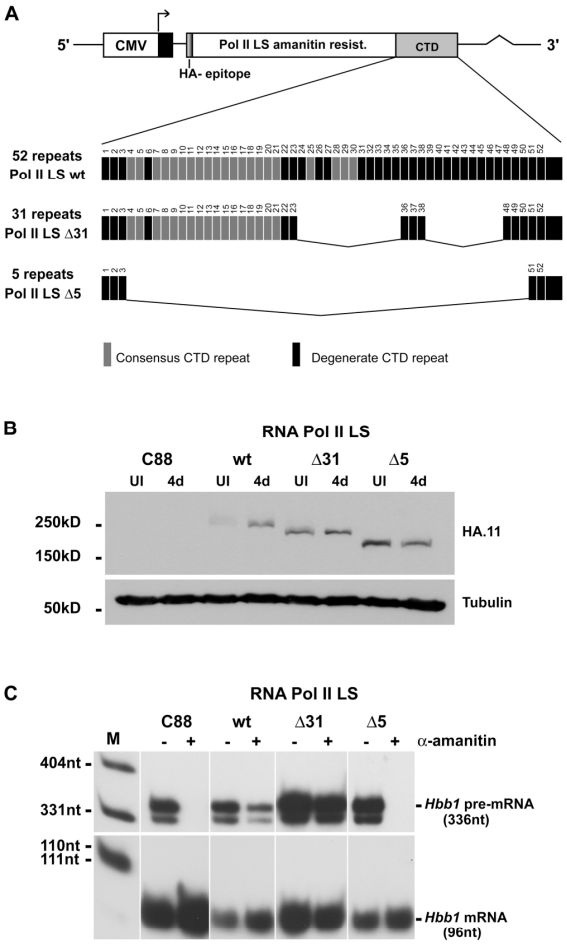


Figure 1. **A large deletion of the CTD abolishes LCR-dependent *Hbb1* transcriptional activation.** (A) Schematic representation of the RNA Pol II LS constructs. (B) Western blotting analysis. Total protein extracts were prepared from untransfected MEL C88 and from cells transfected with the indicated constructs, both before (UI) and after 4 d of erythroid differentiation (4d). The blot was incubated with antibodies anti-HA and anti-tubulin. (C) S1-nuclease protection assay. MEL C88 cells and cells transfected with the indicated constructs were induced for 4 d either with (+) or without (−) treatment with α-amanitin for 17 h. *Hbb1* expression was analyzed with a probe that produces protected fragments of 336 and 96 nt for the pre-mRNA and mRNA, respectively. Due to its long half-life, high levels of mRNA synthesized before treatment with α-amanitin persist in the treated cells.

that transcription by the endogenous RNA Pol II was abolished (Fig. 1 C, lane C88 +). In marked contrast, *Hbb1* pre-mRNA was present in clones expressing the α-amanitin–resistant forms of RNA Pol II LS containing either the full-length CTD (Fig. 1 C, lane wt +) or the Δ31 truncation (Fig. 1 C, lane Δ31 +), but not the Δ5 variant (Fig. 1 C, lane Δ5 +). The finding that none of the clones expressing RNA Pol II Δ5 were able to support transcription of *Hbb1* was surprising, taking into account that this mutant was previously shown to transcribe *HBB* under control of the SV40 promoter (McCracken et al., 1997b) and a rat homeobox reporter gene stably integrated into the genome of

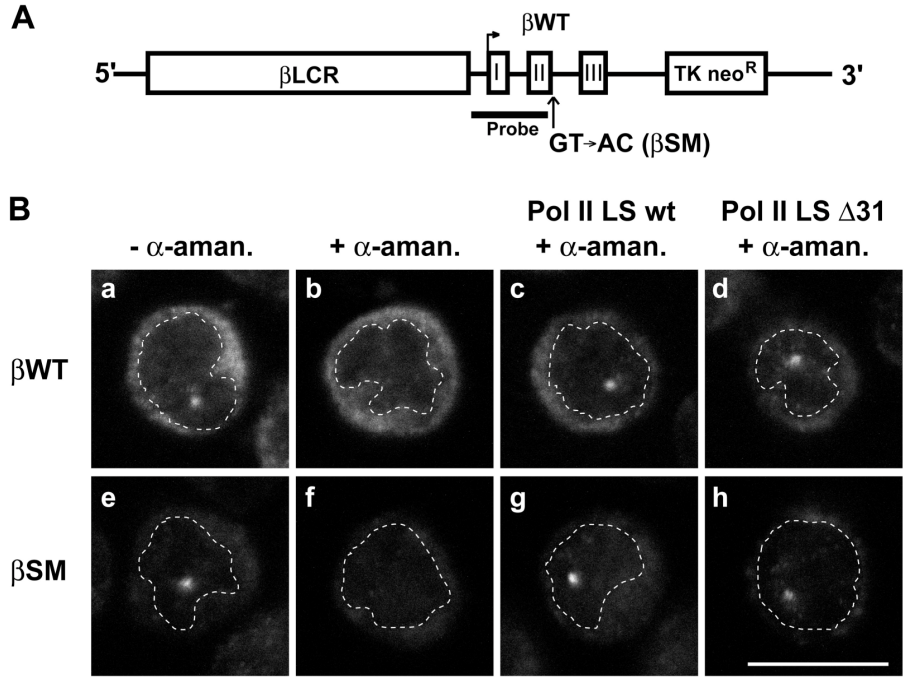


Figure 2. **Visualization of nascent RNA transcribed by transgene-derived RNA Pol II LS.** (A) Schematic representation of wild-type (βWT) and splice mutant (βSM) human β-globin (*HBB*) constructs. Exons in the *HBB* gene are numerated (I, II, III). The probe used for FISH is complementary to exon I, intron I, and 205 nt of exon II. (B) Detection by FISH of βWT and splice mutant βSM RNA synthesized by either endogenous RNA Pol II or transgene-derived constructs, as indicated. All cells were induced for 3 d and either treated or untreated with α-amanitin as indicated. Dashed lines indicate the periphery of the nucleus. Bar, 10 μm.

HeLa cells (Misteli and Spector, 1999). Additional studies indicated that the truncated variant of the CTD with only 5 heptad repeats did not affect TATA-box–mediated transcription (Gerber et al., 1995; McCracken et al., 1997b). However, this same form of the CTD in yeast and humans abolished activator-dependent induction of transcription of specific genes (Allison and Ingles, 1989; Gerber et al., 1995; Meininghaus and Eick, 1999). Furthermore, nuclear run-on experiments in mammalian cells suggested a global defect in transcription of endogenous genes (Meininghaus et al., 2000). Because endogenous murine *Hbb1* is under control of the LCR (Collis et al., 1990), our observation that RNA Pol II Δ5 fails to support LCR-dependent transcription is consistent with previous data indicating a requirement of the CTD for enhancer-driven transcription (Allison and Ingles, 1989; Gerber et al., 1995; Meininghaus and Eick, 1999).

Deletion of the CTD to 31 heptads causes mRNA retention at the site of transcription

Having selected MEL cell clones that express functional α-amanitin–resistant wild-type or Δ31 RNA Pol II LS, we next super-transfected these cells with an *HBB* transgene that was either wild type (βWT) or a mutant variant possessing a GT to AC mutation at the 5′ splice site of the second intron (βSM) (Fig. 2 A). As this mutation inhibits splicing and causes retention of the RNA at the transcription site (Custódio et al., 1999), we were interested in determining whether the CTD is involved in recognition of the resulting aberrantly processed βSM pre-mRNA. Cells were induced for 3 d and analyzed by FISH. As previously described (Custódio et al., 1999), the transcription site of the βWT or βSM transgene is detected as a focus in the nucleus (Fig. 2 B). The wild-type mRNA is exported from the nucleus and accumulates in the cytoplasm (Fig. 2 B, a), in contrast to the mutant RNA, which is not detected in the cytoplasm (Fig. 2 B, e). Treatment of cells devoid of exogenous transgene-derived RNA Pol II with α-amanitin results in the disappearance of nuclear foci (Fig. 2 B, b and f). Contrastingly, nuclear foci remain clearly visible in cells transfected with α-amanitin–resistant forms of RNA Pol II LS (Fig. 2 B, c, d, g, and h).

We have previously shown that treating MEL cells with the transcription inhibitor actinomycin D for a short period of time (5–15 min) causes a dramatic, rapid reduction in the relative number of cells that contain a detectable nuclear focus of βWT transcripts within the nucleus, whereas the percentage of cells harboring βSM RNA foci remained largely unaltered, suggesting that these mutant RNAs were not being released from the site of transcription (Custódio et al., 1999). We therefore performed the same assay using cells that express the α-amanitin–resistant forms of RNA Pol II LS in an effort to gain insight into the possible role of the CTD in the process of mRNA release from the transcription site. The results show that actinomycin D treatment of cells expressing either endogenous RNA Pol II LS or the α-amanitin–resistant full-length CTD transgene product resulted in a significant decrease in the percentage of cells with βWT

transcription foci (Fig. 3 A, a–d; and Fig. 3 C, a), as well as in the intensity of the remaining signal (Fig. 3 D). However, when transcription is dependent on the α-amanitin–resistant RNA Pol II LS containing the truncated Δ31 CTD, actinomycin D treatment does not cause any significant reduction in the percentage of cells with a visible focus (Fig. 3 A, e, f; and Fig. 3 C, a) or in the mean fluorescence intensity of each focus (Fig. 3 D). We therefore conclude that the transcripts are not being efficiently released. Parallel experiments performed with cells expressing the mutant *HBB* transgene showed that treatment with actinomycin D causes no significant change in the percentage of cells with a visible nuclear focus of βSM RNA (Fig. 3 B), regardless of whether this gene is transcribed by full-length (Fig. 3 B c, d; and Fig. 3 C, b) or truncated (Fig. 3 B, e, f; and Fig. 3 C, b) CTD versions of α-amanitin–resistant RNA Pol II LS. Thus, reducing the CTD to 31 heptad repeats is sufficient to prevent release of RNA transcribed from a normal gene while it does not interfere with the ability to retain transcripts derived from a gene with a severe splice mutation.

RNA transcribed by RNA Pol II LS Δ31 is spliced, cleaved, and polyadenylated

Previous studies have shown that deletion mutants of the CTD induce defects in splicing and 3′ end cleavage (McCracken et al., 1997b; Fong and Bentley, 2001; Fong et al., 2003). We therefore analyzed wild-type *HBB* (βWT) transcripts produced by endogenous or exogenous transgene RNA Pol II LS by RNase protection assays to monitor splicing of introns 1 and 2 as well as cleavage at the poly(A) addition site (Fig. 4 A). As expected, no signal for unspliced (US; Fig. 4 B, lane 2) and uncleaved (UC; Fig. 4 B, lane 4) *HBB* RNA was detected after inactivation of the endogenous RNA Pol II LS with α-amanitin (Fig. 4 B). However, bands corresponding to spliced (SP) and cleaved (CL) RNA were still present, most probably due to the long half-life of *HBB* mRNA.

To avoid complications caused by the presence of *HBB* mRNA synthesized by the endogenous RNA Pol II LS and which had accumulated before α-amanitin treatment, cells were exposed to this toxin immediately after the first day of induced erythroid differentiation. In MEL cells containing the βWT transgene and which rely exclusively on the endogenous RNA Pol II for transcription, this resulted in massive cell death, while there was survival of cells transfected with the α-amanitin–resistant forms of RNA Pol II LS. In these cells, after inactivation of the endogenous RNA Pol II LS, bands corresponding to both unspliced and spliced (Fig. 4 C, lanes 2 and 3) as well as uncleaved and cleaved (Fig. 4 C, lanes 5 and 6) products were clearly detected. In a control experiment, we analyzed RNA extracted from MEL cells containing the βWT transgene that had undergone the same period (3 d) of differentiation but not exposed to α-amanitin (Fig. 4 C, lanes 1 and 4). The results (Fig. 4 D) revealed no substantial reduction in the percentage of spliced and 3′ cleaved βWT mRNA transcribed by RNA Pol II LS Δ31 compared with that synthesized by either RNA Pol II LS wild type or endogenous RNA Pol II. We further observed that poly(A) tail length of *HBB* mRNA was similar in transcripts synthesized by either endogenous RNA Pol II RNA (Fig. 4 E, lane 2), RNA Pol II LS wild type (Fig. 4 E, lane 3), or Pol II LS Δ31 (Fig. 4 E, lane 4). Thus, a CTD with heptad repeats 1–23, 36–38, and 48–52 followed by the unique terminal 10-amino acid motif is sufficient to support efficient pre-mRNA processing, as predicted from previous studies (Fong et al., 2003; Rosonina and Blencowe, 2004).

Collectively, our data reveal that βWT transgene mRNA synthesized by the RNA Pol II Δ31 mutant is efficiently spliced, cleaved, and polyadenylated and yet remains in close vicinity to the gene after inhibition of transcription. This strongly suggests that the CTD is required for mRNA release from the site of transcription in a manner independent of splicing and 3′ formation. A possible splicing-independent involvement of the CTD in RNA release was also noted by Bentley and colleagues, who observed that intron-less pre-mRNA synthesized by a terminal 10-amino acid motif mutant CTD remained at the site of transcription, whereas synthesis by a wild-type CTD resulted in RNA release (Bird et al., 2005). However, the use of an intron-less reporter gene in these studies precluded definitive conclusions to be drawn.

EJC proteins and the PM/Scl-100 exosome subunit are recruited to nascent transcripts synthesized by CTD mutant RNA Pol II

We have previously shown that exon junction complex (EJC) proteins and core spliceosome components (U snRNPs) accumulate on nascent wild-type *HBB* transcripts, but fail to associate with mutant transcripts that are not released from the transcription site (Custódio et al., 2004). To assess if the presence of truncation mutants of the RNA Pol II LS CTD affected this process, we conducted double-labeling (FISH plus immunocytochemical staining) of MEL cells that contained the βWT transgene transcribed by either endogenous or exogenous RNA Pol II (full-length or the Δ31 CTD mutant) after treatment with α-amanitin as before (Fig. 5). We used a probe to detect βWT transcripts and antibodies to detect snRNP Sm proteins (Fig. 5, a–c), and EJC components SRm160 (Fig. 5, e–g) and Aly/REF (Fig. 5, i–k). Using a previously described quantitative single-cell assay (Misteli et al., 1998; Mabon and Misteli, 2005), we detected all three proteins recruited to nascent transcripts irrespective of CTD length (Fig. 5, a′–k′ and d, h, and l). Moreover, we observed that colocalization of EJC proteins with transcripts synthesized by RNA Pol II LS Δ31 CTD persists after inhibition of transcription by actinomycin D, which adds additional evidence that RNA transcribed by RNA Pol II harboring a truncated Δ31 CTD and retained at the transcription site is normally spliced.

In the yeast *Saccharomyces cerevisiae*, retention of defective mRNA at the site of transcription requires Rrp6p and other components of the nuclear exosome, suggesting that this complex is part of a quality control checkpoint that monitors for correct processing of pre-mRNA (Hilleren et al., 2001; Jensen et al., 2001; Libri et al., 2002; Dunn et al., 2005). We therefore investigated whether the mammalian orthologue of Rrp6p associates with nascent βWT transcripts by immunofluorescence using a specific antibody (Brouwer et al., 2001). This protein was readily detectable throughout the nucleoplasm with a high

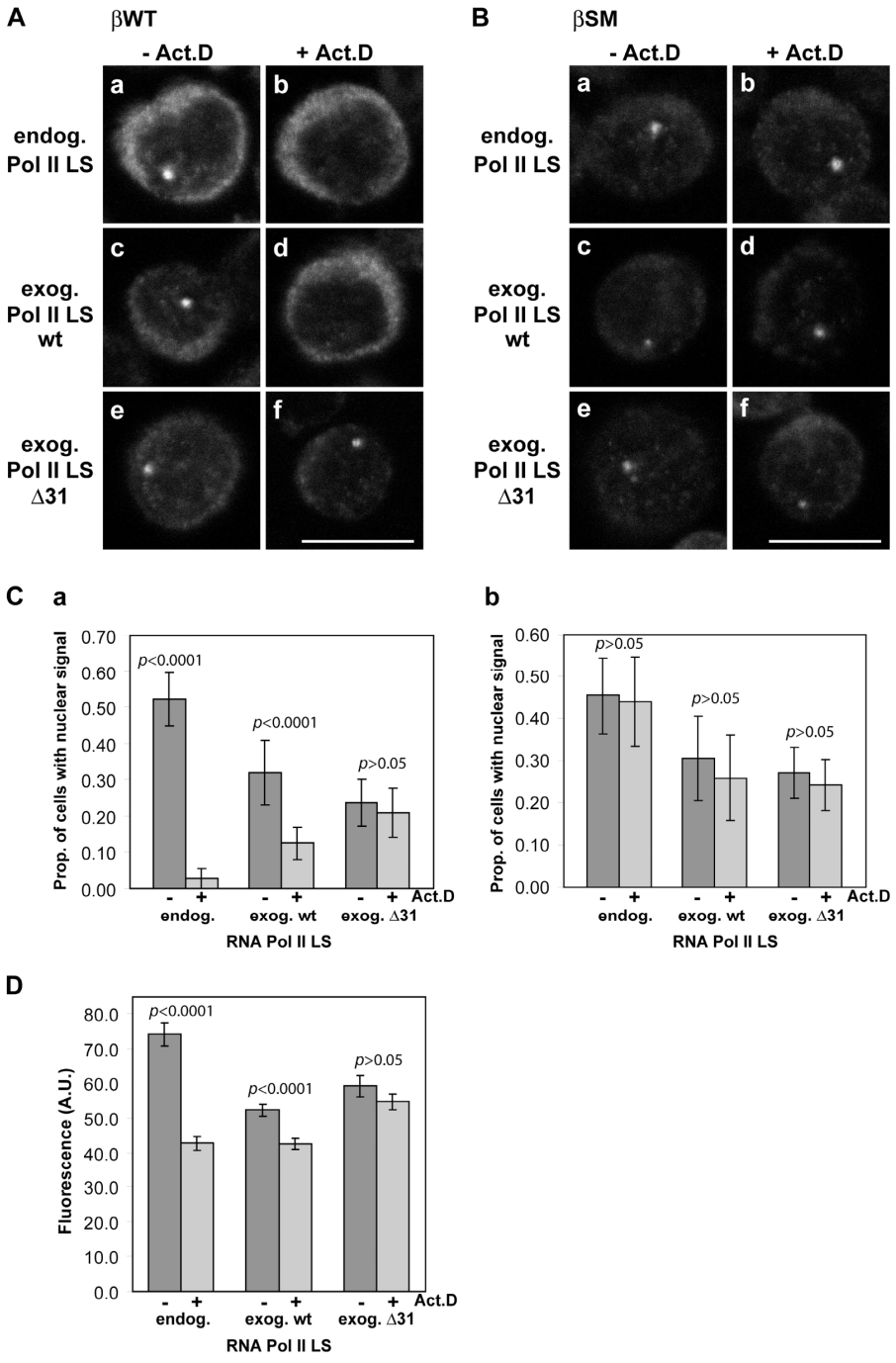


Figure 3. **The CTD Δ31 mutation causes mRNA retention at the site of transcription.** RNA transcribed from βWT (A) and splice mutant βSM (B) transgenes was visualized by FISH. Transcription was by endogenous or exogenous RNA Pol II LS, as indicated. Treatment with actinomycin D was for 15 min. Bar, 10 μm. (C) The proportion of cells with a nuclear RNA focus, before and after actinomycin D. A total of 500–700 cells were counted in each experiment. Results are presented as means ± SD for at least three independent experiments; P values relative to nontreated cells (*t* test) are indicated. Due to a combination of the asynchronous nature of the cell cultures and position-effect variegation, not all cells contain a visible nuclear focus. As the fraction of cells with a nuclear focus varies between cell lines, the interpretation of the data relies on the comparison of the fraction of cells from the same line that contain a nuclear focus before and after treatment with actinomycin D. (D) Mean fluorescence intensity (arbitrary units, AU) of the nuclear βWT RNA focus, before and after actinomycin D. A total of 100 nuclear RNA foci were quantified in each experiment. Results are presented as mean fluorescence intensity ± SE for two independent experiments; P values relative to nontreated cells (*t* test) are indicated.

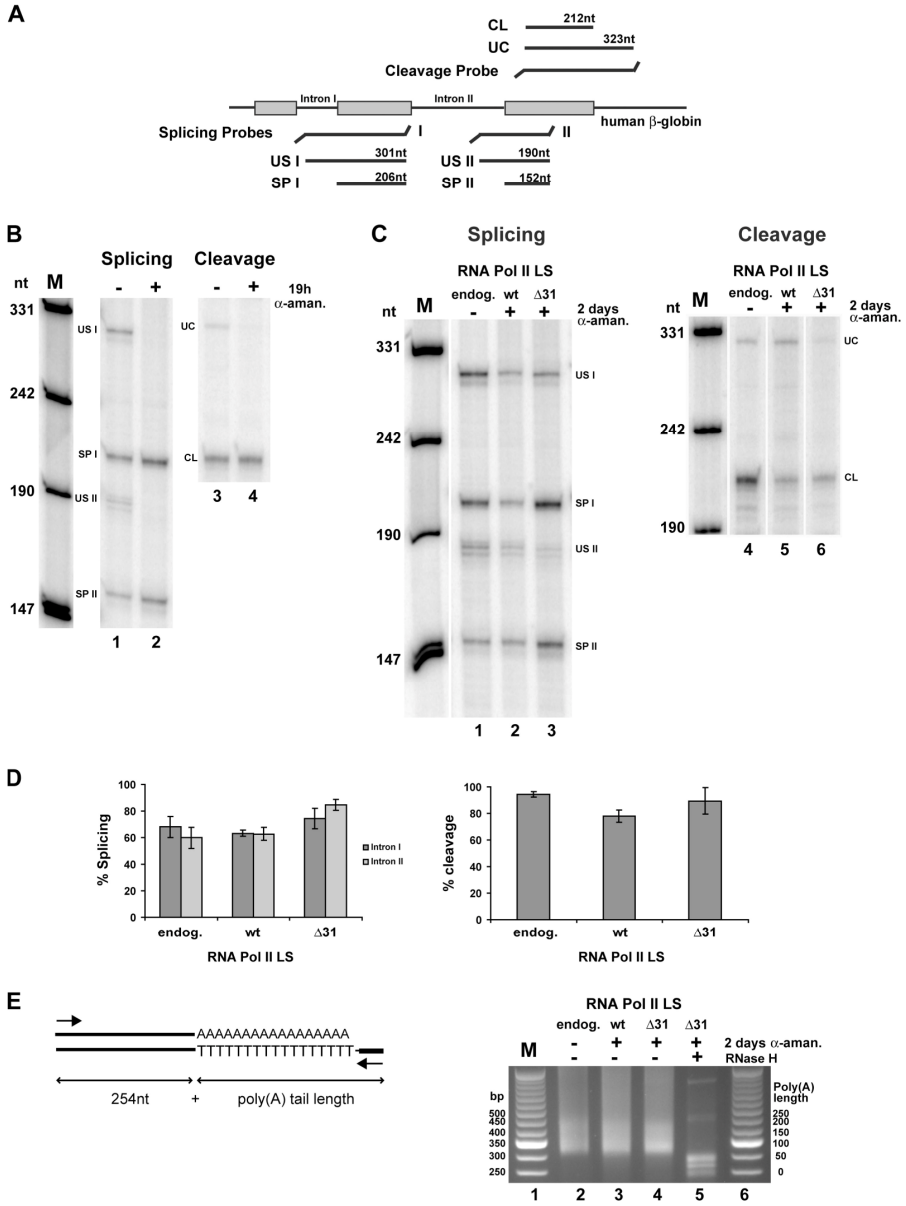


Figure 4. **The CTD Δ31 mutant supports splicing, cleavage, and polyadenylation.** (A) Schematic representation of the RNase protection assay probes. (B) Nuclear RNA (2 μg) from MEL cells transfected with βWT (Fig. 2) induced to differentiate for 3 d and untreated (−) or treated (+) with α-amanitin for 19 h were analyzed by an RNase protection assay using the indicated probes; intron I unspliced (US I), intron II unspliced (US II), intron I spliced (SP I), intron II spliced (SP II), uncleaved (UC), and cleaved (CL). (C) Nuclear RNA fractions from cells transcribing the βWT transgene either by endogenous Pol II LS (endog.; 2.5 μg) or by exogenous RNA Pol II LS wild-type (wt; 5 μg) or RNA Pol II LS Δ31 (Δ31; 5 μg) that were induced to differentiate for 3 d and untreated (−) or treated (+) with α-amanitin for 2 d were analyzed using the same probes. (D) Quantification of splicing and cleavage efficiencies. The amount of unspliced (US), spliced (SP), uncleaved (UC), and cleaved (CL) RNA from three independent experiments was determined from PhosphorImager data corrected for [^{32}P]U content of the RNase-protected fragment. Percentages of splicing and cleavage were calculated by dividing the value of spliced or cleaved product by the sum of the values of spliced and unspliced products or cleaved and uncleaved products, respectively. (E) Poly(A) tail length analysis of *HBB* transcripts. The diagram on the left illustrates the PCR-amplified products. Primers were designed to amplify a 254-nt product if the mRNA is not polyadenylated and any length over 254 nt is contributed by the poly(A) tail. On the right, an ethidium bromide–stained agarose gel with the products obtained from cells transcribing the βWT transgene either by endogenous Pol II LS (endog.; lane 2) or by exogenous RNA Pol II LS wild-type (wt; lane 3) or RNA Pol II LS Δ31 (Δ31; lanes 4 and 5) that were induced to differentiate for 3 d and untreated (−) or treated (+) with α-amanitin for 2 d. RNA from lane 5 was treated with RNase H in the presence of oligo $dT_{12–18}$. Lanes 1 and 6 contain a 50-bp ladder. Similar results were obtained in two independent experiments.

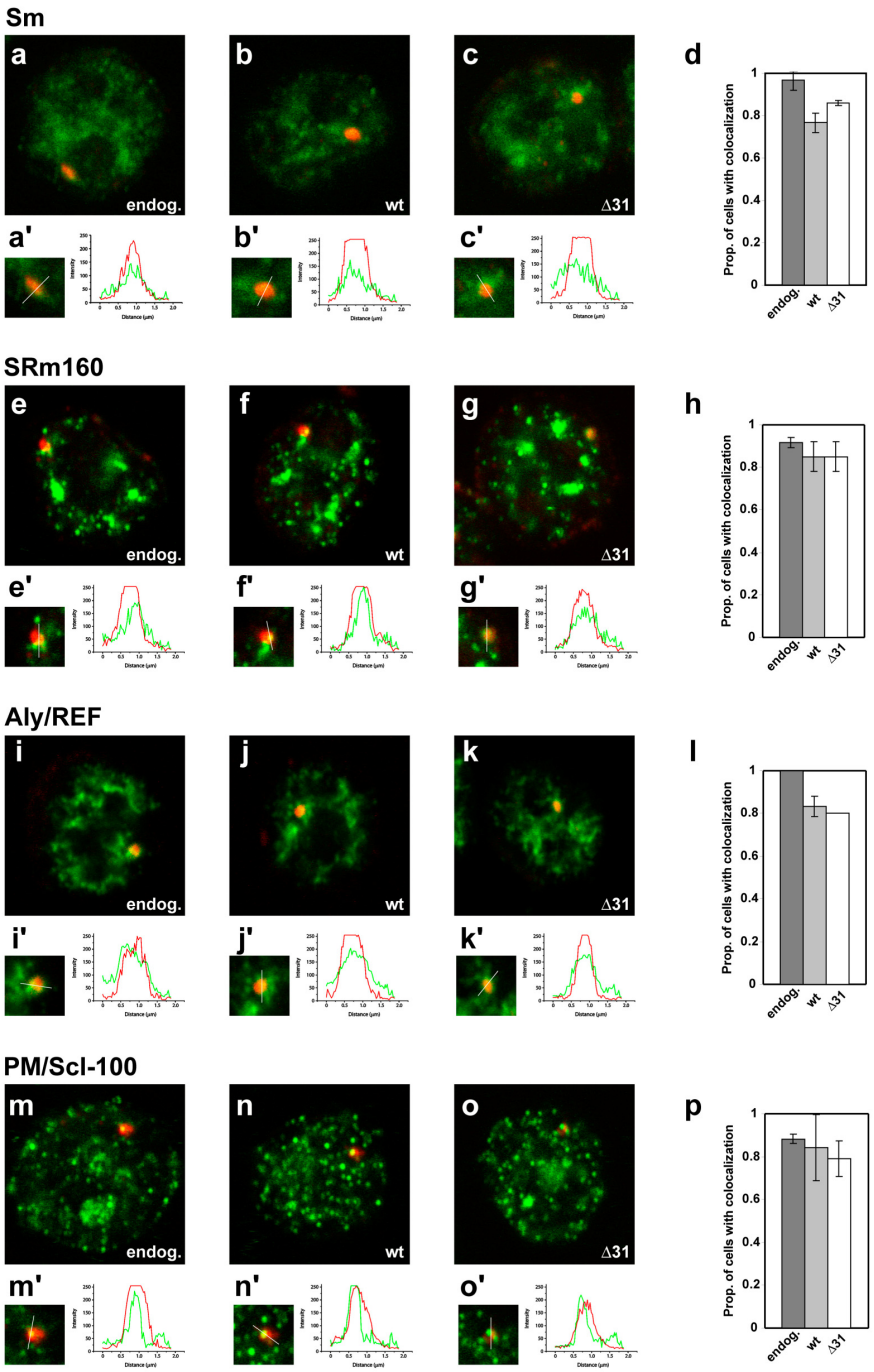


Figure 5. **EJC proteins and the exosome subunit PM/Scl-100 are recruited to the *HBB* transcription site.** MEL cells were induced for 3 d and treated with α-amanitin. The βWT transgene (Fig. 2) was transcribed by endogenous RNA Pol II (a, e, i, and m), RNA Pol II LS wt (b, f, j, and n), or RNA Pol II LS Δ31 (c, g, k, and o). Cells were double-labeled for *HBB* RNA (red staining) and for protein (green staining) using the indicated antibodies. Relative fluorescence intensity (in arbitrary units) was measured along a line across the βWT RNA focus. We considered a positive recruitment when the intensity of the protein signal at the focus was more than twofold higher than the global nucleoplasmic signal (Mabon and Misteli, 2005). Representative results are shown for the indicated proteins. The proportion of cells exhibiting recruitment was quantified. Values represent average from 45 transcription sites analyzed in two independent experiments ± SD.

concentration at the site of βWT transgene transcription (Fig. 5, m–o). Similar results were observed in cells that express exogenous RNA Pol II LS with either a wild-type (Fig. 5, n, n′, and p) or truncated Δ31 (Fig. 5, o, o′, and p) CTD.

Collectively, these data indicate that recruitment of neither EJC proteins nor nuclear exosome Rrp6 class of proteins to nascent mRNA is sufficient for its release from the site of transcription.

Conclusions

Accumulation of nascent mRNA in close proximity to their transcription site is thought to represent a surveillance mechanism that prevents defectively processed transcripts from entering the flow to the cytoplasm (Saguez et al., 2005; Gorski et al., 2006). A major player in cotranscriptional pre-mRNA maturation is the CTD of RNA Pol II, which acts by facilitating specific interactions between processing factors while the transcript is still attached to the polymerase (for review see Bentley, 2005). Previous work revealed that a mutation of the terminal 10-amino acid motif of the CTD inhibited splicing, 3′ end cleavage (Fong et al., 2003), and RNA release from the site of transcription (Bird et al., 2005). Based on these observations it was proposed that the CTD is required for transcript release as a consequence of its role in splicing and 3′ end cleavage (Bird et al., 2005). However, in this report we show that a partial truncation of the CTD (Δ31) containing heptads 1–23, 36–38, and 48–52 including the terminal 10-amino acid motif is sufficient to support transcription, splicing, 3′ end cleavage, and polyadenylation, but the newly synthesized mRNA fails to be efficiently released (Figs. 3 and 4). These novel observations imply that the CTD is involved in processes that control the release of transcripts by a mechanism independent from splicing and cleavage. Collectively with previously reported data (Bird et al., 2005), our results further suggest that different segments of the CTD play distinct roles in pre-mRNA processing and mRNA release from the vicinity of the gene template.

Although the mechanism via which the RNA Pol II LS CTD is involved in the release of mRNA from the transcription site is unknown, we speculate that the Δ31 CTD truncation mutant used in this study fails to bind and therefore recruit protein factors required to complete maturation of spliced and 3′ end cleaved/polyadenylated mRNA into export-competent mRNPs. We further propose that this defect in production of fully mature mRNA activates a quality control checkpoint or surveillance mechanism that prevents diffusion of mRNPs to the nuclear pores by tethering them near the gene template, with stalled mRNA being subsequently degraded by the exosome present at the transcription site. A prediction from this mechanistic model is that most mRNA transcribed by the Δ31 CTD mutant should not translate into protein. In good agreement with that prediction is the observation that the Δ31 CTD mutation cannot support long-term cell viability (Meininghaus et al., 2000).

Materials and methods

MEL cell culture and stable transfections

The maintenance, induction of erythroid differentiation, and stable transfection of the MEL cell line C88 were described elsewhere in detail (Antoniou, 1991). The cells were cotransfected as previously described (Custódio et al., 2006) with an α-amanitin–resistant RNA Pol II LS gene with either a full-length CTD (52 heptad repeats, wild type) or a CTD with 31 (Δ31) or 5 (Δ5) repeats (Fig. 1 A; Bartolomei et al., 1988; Gerber et al., 1995; plasmids provided by W. Schaffner, University of Zürich, Zürich, Switzerland) and a plasmid containing the βLCR (Collis et al., 1990) modified to carry a puromycin resistance gene under the control of a phosphoglycerate kinase promoter (Millevoi et al., 2002). Before transfection the RNA Pol II LS wild-type plasmid was linearized with MluI, the Δ31 and Δ5 plasmids were linearized with ClaI and the micro-βLCR plasmid was linearized with PvuI. Stable transfected clones were obtained by culture in the presence of 2.5 μg/ml puromycin (Sigma-Aldrich). Clones selected for further studies were then super-transfected with either the wild-type (βWT) or mutant (βSM) *HBB* genes in the micro-βLCR expression vector (Fig. 2 A; Custódio et al., 1999) with stably transfected cells isolated in the presence of 800 μg/ml G418. The function of endogenous RNA Pol II LS was inhibited by adding α-amanitin (Sigma-Aldrich) to the cell culture medium to a final concentration of 2.5 μg/ml.

In situ hybridization and immunofluorescence analysis

HBB transcripts were visualized by FISH (Custódio et al., 1999) and double labeling for RNA and protein was as previously described (Custódio et al., 2004). The probe used for FISH was a 740-bp fragment of the human β-globin gene extending from the SnaBI site at −265 bp from the transcriptional start point to the BamHI site at +475 bp. The fragment was labeled by nick-translation with either digoxigenin-11-dUTP (Roche) or Cy3-AP3-dUTP (GE Healthcare). The following primary antibodies were used for immunofluorescence: rabbit polyclonal directed against SRm160 (1:500; Blencowe et al., 1998; provided by B. Blencowe, University of Toronto, Ontario, Canada), mouse monoclonal directed against Aly/REF (1:100; clone 11G5; AbCam), human autoantiserum C45, specific for Sm proteins (1:75; provided by W. van Venrooij, University of Nijmegen, Nijmegen, Netherlands) and rabbit serum against PM/Scl-100 (1:75; provided by Ger Pruijn, University of Nijmegen). The secondary antibodies used were: AlexaFluor 488–conjugated goat anti–rabbit IgG (1:200; Jackson ImmunoResearch Laboratories, Inc.), AlexaFluor 488–conjugated goat anti–mouse IgG (1:200; Jackson ImmunoResearch Laboratories, Inc.), and FITC-conjugated donkey anti–human IgG (1:100; Jackson ImmunoResearch Laboratories, Inc.).

Microscopy and image quantification

Images were acquired on a laser scanning confocal microscope (LSM 510 or LSM 510 META; Carl Zeiss MicroImaging, Inc.) using the PlanApochromat 63×/1.4 objective. FITC and AlexaFluor 488 fluorescence was detected using the 488-nm line of the argon ion laser. Cy3 was excited with the 543-nm line of the helium-neon laser on the Zeiss LSM 510 and with the DPSS 561-10 laser on the Zeiss LSM 510 META confocal microscope. To quantify the intensity of the nuclear foci, single-cell images were acquired with no saturated pixels, always using the same settings. The mean intensity of fluorescence in the nuclear RNA focus was determined using ImageJ (http://rsb.info.nih.gov/ij/). Line profiles were obtained from unprocessed images using the LSM 510 software.

Western blotting

Total cell extracts were prepared as described in Custódio et al. (2006). Volumes of total extract equivalent to 10^6 cells were fractionated on a 7% polyacrylamide-SDS gel and proteins were transferred to nitrocellulose in 24 mM Tris, 193 mM glycine, and 20% methanol for 16 h at 30 mA. Western blotting with mouse monoclonal against HA epitope (HA.11; Covance) and mouse monoclonal against α-tubunin (clone B-5-1-2; Sigma-Aldrich) was as previously described (Custódio et al., 2006).

S1-nuclease and RNase protection assays

The levels of endogenous murine β^{major}-globin gene (*Hbb1*) pre-mRNA were analyzed by an S1-nuclease protection assay as described (Custódio et al., 2006). RNase protection assays were performed as described (McCracken et al., 2002; Custódio et al., 2004). *HBB* RNase protection probes (McCracken et al., 1997a) were prepared by in vitro transcription with T7 RNA polymerase in the presence of α-[^{32}P]UTP (GE Healthcare). In each reaction nuclear RNA, prepared as described (Custódio et al., 2004), was incubated with excess of the antisense RNA probe overnight at 50°C, and the hybridization products were digested with a mixture of RNase T1 and A (Ambion) at 37°C for 1 h. The RNase-protected fragments were resolved on a 6% denaturing polyacrylamide gel and the intensity of the bands was quantified using a PhosphorImager (Molecular Dynamics). After quantification of each gel band, background was subtracted and the values were normalized for different U residue contents of the protected probe fragments.

Poly(A) tail length analysis

The poly(A) tail length analysis of *HBB* transcripts was performed by PCR using the ligase-mediated poly(A) test (LM-PAT) described by Sallés et al. (1999). For cDNA synthesis we used 3 μg of total RNA, 50 ng of phosphorylated oligo dT_{12}, and 1 μg of oligo(dT)-anchor primer. As a control, RNA was deadenylated by digestion with RNase H in the presence of oligo $dT_{12–18}$. For PCR amplification, 1 μl of the template LM-PAT cDNA was added to a standard 25-μl PCR reaction containing 12.5 pmol of oligo(dT)-anchor primer and 12.5 pmol of an *HBB* specific primer that hybridizes 254 nt upstream from the polyadenylation signal (5′-GCAACGTGCTGGTCTGTGTGCTG-3′). The amplified products were resolved by electrophoresis on a 2% agarose gel stained with ethidium bromide.

We are grateful to Elmar Wahle for advice on the polyadenylation assay. We further thank the following groups for generously providing reagents: W. Schaffner for the CTD constructs; W. Van Venrooij for the human autoantiserum C45; B. Blencowe for the antibody against SRm160; and Ger Pruijn for the rabbit serum against PM-Scl100.

This work was supported by Fundação para a Ciência e Tecnologia, Portugal, the European Science Foundation (EuroDYNA Programme), and the European Commission (EURASNET, LSHG-CT-2005-518238).

Submitted: 19 December 2006
Accepted: 22 September 2007

References

Aguilera, A. 2005. Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. *Curr. Opin. Cell Biol.* 17:242–250.

Allison, L.A., and C.J. Ingles. 1989. Mutations in RNA polymerase II enhance or suppress mutations in GAL4. *Proc. Natl. Acad. Sci. USA.* 86:2794–2798.

Antoniou, A. 1991. Induction of erythroid-specific expression in murine erythroleukemia (MEL) cells. *In* Methods in Molecular Biology: Gene Transfer and Expression Protocols. Vol. 7. E.J. Murray, editor. The Human Press Inc., Clinfton, NJ. 421–434.

Antoniou, M., and F. Grosveld. 1990. β-globin dominant control region interacts differently with distal and proximal promoter elements. *Genes Dev.* 4:1007–1013.

Bartolomei, M.S., and J.L. Corden. 1987. Localization of an alpha-amanitin resistance mutation in the gene encoding the largest subunit of mouse RNA polymerase II. *Mol. Cell. Biol.* 7:586–594.

Bartolomei, M.S., N.F. Halden, C.R. Cullen, and J.L. Corden. 1988. Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. *Mol. Cell. Biol.* 8:330–339.

Bentley, D.L. 2005. Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr. Opin. Cell Biol.* 17:251–256.

Bird, G., N. Fong, J.C. Gatlin, S. Farabaugh, and D.L. Bentley. 2005. Ribozyme cleavage reveals connections between mRNA release from the site of transcription and pre-mRNA processing. *Mol. Cell.* 20:747–758.

Blencowe, B.J., R. Issner, J.A. Nickerson, and P.A. Sharp. 1998. A coactivator of pre-mRNA splicing. *Genes Dev.* 12:996–1009.

Blom van Assendelft, G., O. Hanscombe, F. Grosveld, and D.R. Greaves. 1989. The β-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell.* 56:969–977.

Brouwer, R., C. Allmang, R. Raijmakers, Y. van Aarssen, W.V. Egberts, E. Petfalski, W.J. van Venrooij, D. Tollervey, and G.J. Pruijn. 2001. Three novel components of the human exosome. *J. Biol. Chem.* 276:6177–6184.

Chapman, R.D., B. Palancade, A. Lang, O. Bensaude, and D. Eick. 2004. The last CTD repeat of the mammalian RNA polymerase II large subunit is important for its stability. *Nucleic Acids Res.* 32:35–44.

Collis, P., M. Antoniou, and F. Grosveld. 1990. Definition of the minimal requirements within the human β-globin gene and the dominant control region for high level expression. *EMBO J.* 9:233–240.

Custódio, N., M. Carmo-Fonseca, F. Geraghty, H.S. Pereira, F. Grosveld, and M. Antoniou. 1999. Inefficient processing impairs release of RNA from the site of transcription. *EMBO J.* 18:2855–2866.

Custódio, N., C. Carvalho, I. Condado, M. Antoniou, B.J. Blencowe, and M. Carmo-Fonseca. 2004. In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei. *RNA.* 10:622–633.

Custódio, N., M. Antoniou, and M. Carmo-Fonseca. 2006. Abundance of the largest subunit of RNA polymerase II in the nucleus is regulated by nucleo-cytoplasmic shuttling. *Exp. Cell Res.* 312:2557–2567.

Dunn, E.F., C.M. Hammell, C.A. Hodge, and C.N. Cole. 2005. Yeast poly(A)-binding protein, Pab1, and PAN, a poly(A) nuclease complex recruited by Pab1, connect mRNA biogenesis to export. *Genes Dev.* 19:90–103.

Fong, N., and D.L. Bentley. 2001. Capping, splicing, and 3′ processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev.* 15:1783–1795.

Fong, N., G. Bird, M. Vigneron, and D.L. Bentley. 2003. A 10 residue motif at the C-terminus of the RNA pol II CTD is required for transcription, splicing and 3′ end processing. *EMBO J.* 22:4274–4282.

Galy, V., O. Gadal, M. Fromont-Racine, A. Romano, A. Jacquier, and U. Nehrbass. 2004. Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell.* 116:63–73.

Gerber, H.P., M. Hagmann, K. Seipel, O. Georgiev, M.A. West, Y. Litingtung, W. Schaffner, and J.L. Corden. 1995. RNA polymerase II C-terminal domain required for enhancer-driven transcription. *Nature.* 374:660–662.

Gorski, S.A., M. Dundr, and T. Misteli. 2006. The road much traveled: trafficking in the cell nucleus. *Curr. Opin. Cell Biol.* 18:284–290.

Hilleren, P., T. McCarthy, M. Rosbash, R. Parker, and T.H. Jensen. 2001. Quality control of mRNA 3′-end processing is linked to the nuclear exosome. *Nature.* 413:538–542.

Jensen, T.H., K. Patricio, T. McCarthy, and M. Rosbash. 2001. A block to mRNA nuclear export in S. cerevisiae leads to hyperadenylation of transcripts that accumulate at the site of transcription. *Mol. Cell.* 7:887–898.

Jensen, T.H., K. Dower, D. Libri, and M. Rosbash. 2003. Early formation of mRNP: license for export or quality control? *Mol. Cell.* 11:1129–1138.

Libri, D., K. Dower, J. Boulay, R. Thomsen, M. Rosbash, and T.H. Jensen. 2002. Interactions between mRNA export commitment, 3′-end quality control, and nuclear degradation. *Mol. Cell. Biol.* 22:8254–8266.

Mabon, S.A., and T. Misteli. 2005. Differential recruitment of pre-mRNA splicing factors to alternatively spliced transcripts in vivo. *PLoS Biol.* 3:e374.

McCracken, S., N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, S. Foster, S. Shuman, and D.L. Bentley. 1997a. 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev.* 11:3306–3318.

McCracken, S., N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, S.D. Patterson, M. Wickens, and D.L. Bentley. 1997b. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature.* 385:357–361.

McCracken, S., M. Lambermon, and B.J. Blencowe. 2002. SRm160 splicing coactivator promotes transcript 3′-end cleavage. *Mol. Cell. Biol.* 22:148–160.

Meininghaus, M., and D. Eick. 1999. Requirement of the carboxy-terminal domain of RNA polymerase II for the transcriptional activation of chromosomal c-fos and hsp70A genes. *FEBS Lett.* 446:173–176.

Meininghaus, M., R.D. Chapman, M. Horndasch, and D. Eick. 2000. Conditional expression of RNA polymerase II in mammalian cells. Deletion of the carboxyl-terminal domain of the large subunit affects early steps in transcription. *J. Biol. Chem.* 275:24375–24382.

Millevoi, S., F. Geraghty, B. Idowu, J.L. Tam, M. Antoniou, and S. Vagner. 2002. A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3′-end processing. *EMBO Rep.* 3:869–874.

Misteli, T., and D.L. Spector. 1999. RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol. Cell.* 3:697–705.

Misteli, T., J.F. Caceres, J.Q. Clement, A.R. Krainer, M.F. Wilkinson, and D.L. Spector. 1998. Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J. Cell Biol.* 143:297–307.

Moore, M.J. 2005. From birth to death: the complex lives of eukaryotic mRNAs. *Science.* 309:1514–1518.

Nguyen, V.T., F. Giannoni, M.F. Dubois, S.J. Seo, M. Vigneron, C. Kedinger, and O. Bensaude. 1996. In vivo degradation of RNA polymerase II largest subunit triggered by alpha-amanitin. *Nucleic Acids Res.* 24:2924–2929.

Rosonina, E., and B.J. Blencowe. 2004. Analysis of the requirement for RNA polymerase II CTD heptapeptide repeats in pre-mRNA splicing and 3′-end cleavage. *RNA.* 10:581–589.

Saguez, C., J.R. Olesen, and T.H. Jensen. 2005. Formation of export-competent mRNP: escaping nuclear destruction. *Curr. Opin. Cell Biol.* 17:287–293.

Sallés, F.J., W.G. Richards, and S. Strickland. 1999. Assaying the polyadenylation state of mRNAs. *Methods.* 17:38–45.

Stiller, J.W., and B.D. Hall. 2002. Evolution of the RNA polymerase II C-terminal domain. *Proc. Natl. Acad. Sci. USA.* 99:6091–6096.

4. Abundance of the largest subunit of RNA polymerase II in the nucleus is regulated by nucleo-cytoplasmic shuttling

To test the hypothesis that the transcription site retention of *HBB* pre-mRNA processing mutants could be mediated by the CTD of the largest subunit of RNA Pol II we established murine cell clones that express a α-amanitin resistant form of this subunit. The cell lines generated were screened for survival and the ability to transcribe in the presence of α-amanitin. Two cell lines that pass this screening process were chosen for further characterisation, one with low levels of expression of the exogenous RNA Pol II LS and another one with very high expression levels of this subunit. These cell lines were used as a model to study the sub-cellular localisation of the largest subunit of RNA Pol II. From this study we concluded that the over-expressed RNA Pol II LS is predominantly hypophosphorylated, soluble and accumulates in the cytoplasm in a CRM1-dependent manner. In contrast the transcriptionally active form of RNA Pol II LS (containing phosphoserine in position 2 of the CTD repeats) is restricted to the nucleus and its levels remain remarkably constant, even in the cell line with very high expression levels of this subunit. These results suggest that the nuclear-cytoplasmic distribution of RNA Pol II LS may be regulated by shuttling and we propose that this may provide a mechanism to control the pool of RNA polymerase subunits that is accessible for assembly of a functional enzyme in the nucleus.

Work published in: Exp. Cell Res. Vol.312 No.13 pp.2557-2567, 2006

Concerning this part of the work I would like to stress that Dr. Michael Antoniou gave a valuable contribution in the generation of the clones stably transfected with the α-amanitin resistant form of the RNA polymerase II largest subunit.

EXPERIMENTAL CELL RESEARCH 312 (2006) 2557–2567

Research Article

Abundance of the largest subunit of RNA polymerase II in the nucleus is regulated by nucleo-cytoplasmic shuttling

Noélia Custódio[a], *Michael Antoniou*[b], *Maria Carmo-Fonseca*[a,*]

[a]*Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028 Lisboa, Portugal*
[b]*Department of Medical and Molecular Genetics, King's College London School of Medicine, Guy's Campus, Guy's Hospital, London, SE1 9RT, UK*

ARTICLE INFORMATION

Article Chronology:
Received 16 February 2006
Revised version received
15 April 2006
Accepted 19 April 2006
Available online 9 May 2006

Keywords:
RNA polymerase II
Nuclear export
CRM1
Nucleo-cytoplasmic shuttling

ABSTRACT

Eukaryotic RNA polymerase II is a complex enzyme composed of 12 distinct subunits that is present in cells in low abundance. Transcription of mRNA by RNA polymerase II involves a phosphorylation/dephosphorylation cycle of the *carboxyl*-terminal domain (CTD) of the enzyme's largest subunit. We have generated stable murine cell lines expressing an α-amanitin-resistant form of the largest subunit of RNA polymerase II (RNA Pol II LS). These cells maintained transcriptional activity in the presence of α-amanitin, indicating that the exogenous protein was functional. We observed that over-expressed RNA Pol II LS was predominantly hypophosphorylated, soluble and accumulated in the cytoplasm in a CRM1-dependent manner. Our results further showed that the transcriptionally active form of RNA Pol II LS containing phosphoserine in position 2 of the CTD repeats was restricted to the nucleus and its levels remained remarkably constant. We propose that nucleo-cytoplasmic shuttling of RNA Pol II LS may provide a mechanism to control the pool of RNA polymerase subunits that is accessible for assembly of a functional enzyme in the nucleus.

Introduction

In eukaryotic cells, the transcription of protein-encoding genes is carried out by a large complex composed of RNA polymerase II and several auxiliary factors. RNA polymerase II is composed of 12 distinct subunits forming a complex of more than 500 kDa in molecular mass that is present in cells in low abundance (reviewed in [1]). Efficient transcription of mRNA by RNA polymerase II proceeds through multiple stages (pre-initiation, initiation and elongation) and requires reversible phosphorylation of the *carboxyl*-terminal domain (CTD) of its largest subunit (RNA Pol II LS). The CTD of RNA Pol II LS is composed of tandem repeats of seven amino acids with the consensus sequence Tyr^1–Ser^2–Pro^3–Thr^4–Ser^5–Pro^6–Ser^7 [2,3]. Although this consensus is highly conserved in eukaryotes, the number of repeats varies from 26 in yeast to 52 in mammals. Five out of the seven residues can be phosphorylated, but residues Ser^2 and Ser^5 are the main targets of phosphorylation [4,5]. When RNA polymerase II is recruited to a promoter, the CTD is hypophosphorylated. The CTD later becomes hyperphosphorylated during transcription elongation [6]. Phosphorylation of Ser^5 residues occurs at or near the promoter, whereas Ser^2 residues are phosphorylated as the polymerases move away from the promoter and become engaged in transcript elongation [7–9]. Finally, the enzyme is recycled for new rounds of transcription by CTD phosphatases, the best studied of which is FCP1 [10,11].

* *Corresponding author*. Fax: + 351 217 999 412.
E-mail address: carmo.fonseca@fm.ul.pt (M. Carmo-Fonseca).

0014-4827/$ – see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.yexcr.2006.04.018

Here we provide for the first time evidence that RNA Pol II LS shuttles between the nucleus and the cytoplasm, and we propose that this may constitute a mechanism to control the pool of RNA polymerase subunits in the nucleus.

Materials and methods

MEL cell culture

The maintenance, stable transfection by electroporation and induction of erythroid differentiation with 2% dimethylsulfoxide (DMSO) of the murine erythroleukemia (MEL) cell line C88 were performed as described in detail elsewhere [12]. To inhibit CRM1-mediated nuclear export, we treated cells with 50nM leptomycin B (LMB, Sigma) for 3 or 5 h before processing for immunofluorescence.

Generation of MEL cell clones expressing RNA Pol II LS

The cells were co-transfected with the previously described construct containing an α-amanitin-resistant RNA Pol II LS gene with a full-length CTD (52 heptad repeats) [13] and a plasmid containing the human β-globin micro-locus control region (βLCR) [14] and the puromycin resistance gene. Some modifications were introduced in the RNA Pol II LS construct to allow linearization of the plasmid before transfection, namely, the introduction at the *Cla*I site (MG position 16353) of the human β-globin poly(A) signal followed by a unique *Mlu*I site. For transfection, 1×10^7 cells were electroporated (960 μF, 250 V) in the presence of 50 μg of RNA Pol II LS plasmid linearized with *Mlu*I and 5 μg of the LCR plasmid linearized with *Pvu*I. Clones were isolated in the presence of 2.5 μg/ml puromycin (Sigma). α-Amanitin (Sigma) was added to the medium to a final concentration of 2.5 μg/ml for 17 h to inhibit the endogenous RNA Pol II.

Extraction of RNA and S1-nuclease protection assay

Isolated clones were amplified and screened for exogenous RNA Pol II LS expression by S1-nuclease protection assay. For this assay, total RNA was extracted from the clones by selective precipitation in the presence of 3 M LiCl and 6 M urea [12]. The expression of RNA Pol II LS was analyzed in 40 μg of total RNA using a double stranded, end-labelled DNA probe. This probe, illustrated in Fig. 1A, was prepared by PCR amplification using the RNA Pol II LS plasmid as the template. The sense primer was in the cytomegalovirus (CMV) promoter region (5′-ACC CCA TTG ACG TCA ATG GGA GTT TGT TTT-3′) and the antisense primer in the coding region of the RNA Pol II LS gene (5′-CCA TCT TCC CAG AGG ATC ATC AGC CAA AGG-3′). This probe allows the simultaneous detection of the exogenous and endogenous Pol II LS transcripts as 300nt and 140nt S1-protected products, respectively.

The expression of the murine β^{maj}-globin gene (pre-mRNA) was analyzed by S1-nuclease protection assay in 10 μg of nuclear RNA from 4-day induced cells either untreated or treated with α-amanitin (2.5 μg/ml) for 17 h. For this assay, nuclear RNA fractions were prepared as previously described [15]. The β^{maj} probe is a 700-bp *Hin*dIII–*Nco*I fragment from the 5′-half of this gene [16] that gives a 336nt S1-nuclease protected product for unspliced transcripts.

The products from the S1-nuclease digestion were resolved on a 6% polyacrylamide gel in the presence of 8 M urea. The gel was dried and exposed to Kodak BioMax X-ray film. For quantification of the signal intensities, a Molecular Dynamics PhosphorImager was used.

Antibodies

The following primary antibodies were used. Mouse monoclonals directed against the following: (i) hemagglutinin (HA) epitope (mAb HA.11; Covance); (ii) tubulin (clone B-5-1-2; Sigma); (iii) RNA Pol II LS (clone ARNA-3; Research Diagnostics, Inc.; [17]); (iv) hypophosphorylated epitope on the CTD of RNA Pol II LS (clone 8WG16; Covance; [18]); (v) hyperphosphorylated epitope on the CTD of RNA Pol II LS (clone H5, mouse IgM; Covance; [19,20]); (vi) hyperphosphorylated epitope on the CTD of RNA Pol II LS (clone CC-3; [21,22]); (vii) Sm antigen of snRNPs (mAb Y12; [23]); and (viii) hnRNP C protein (clone 4F4; [24]).

Western blotting

Total protein extracts were prepared according to Ref. [22] with some minor modifications. MEL cells were washed in PBS, and the pellet solubilized with TD buffer [0.5% Triton X-100, 0.5% sodium deoxycholate, 50 mM Tris–HCl, 250 mM NaCl, 5 mM EDTA, 50 mM NaF, 2 mM Na_3VO_4, 5 mM β-glycerophosphate, 0.2 μM Okadaic acid (LC Laboratories) and protease inhibitor cocktail (Complete™, Mini; Boehringer Mannheim); pH 7.5] on ice for 15 min. Insoluble material was centrifuged for 30 min at 13,000 × *g*, 4°C, and the supernatant rapidly frozen on liquid nitrogen and stored at –80°C. For nuclear and cytoplasmic fractionation, cells were washed in PBS, resuspended in 2× packed cell volume of ice-cold buffer A [10 mM HEPES, 1.5 mM $MgCl_2$, 10 mM KCl, 5 mM NaF, 1 mM Na_3VO_4, 0.5 mM DTT, 0.2 μM Okadaic acid and protease inhibitor cocktail (Complete™, Mini, EDTA-free; Boehringer Mannheim); pH 7.5] and incubated on ice for 15 min to allow swelling. Twice the initial volume of buffer A was then added and cells were forced repeatedly (10–12 times) through a 25-gauge needle and centrifuged for 5 min at 960×*g*, 4°C. The supernatant (cytoplasmic fraction) was removed, centrifuged again at full speed for 5 min, 4°C, and aliquots were frozen. The nuclear pellet was solubilized in high salt TD buffer (similar to TD buffer, but with 400 mM NaCl and 10% glycerol) on ice for 15 min and processed as described for the total extract. Before electrophoresis, the extracts were boiled for 5 min in 2× SDS sample buffer (80 mM Tris–HCl, 4.5% SDS, 150 mM DTT, 16% glycerol, 0.01% bromophenol blue; pH 6.8). Volumes of total extract equivalent to 2×10^6 cells and 30 μg of nuclear and cytoplasmic extracts were fractionated on either a 7% or a 10% polyacrylamide–SDS gel. Proteins were transferred to nitrocellulose in 48 mM Tris, 39 mM glycine, 0.04% SDS, 10% methanol using a semi-dry electrophoretic transfer cell (BioRad; 2 mini gels were transferred simultaneously for 1 h: 30 min at 250 mA). Blots were probed with mAbs, HA.11 (1/500), α-tubulin (1/2000), ARNA-3 (1/1000), 8WG16 (1/500), H5 (1/500) and CC-3 (1/1000) for 3 h in 2.5% milk in PBS. For detection,

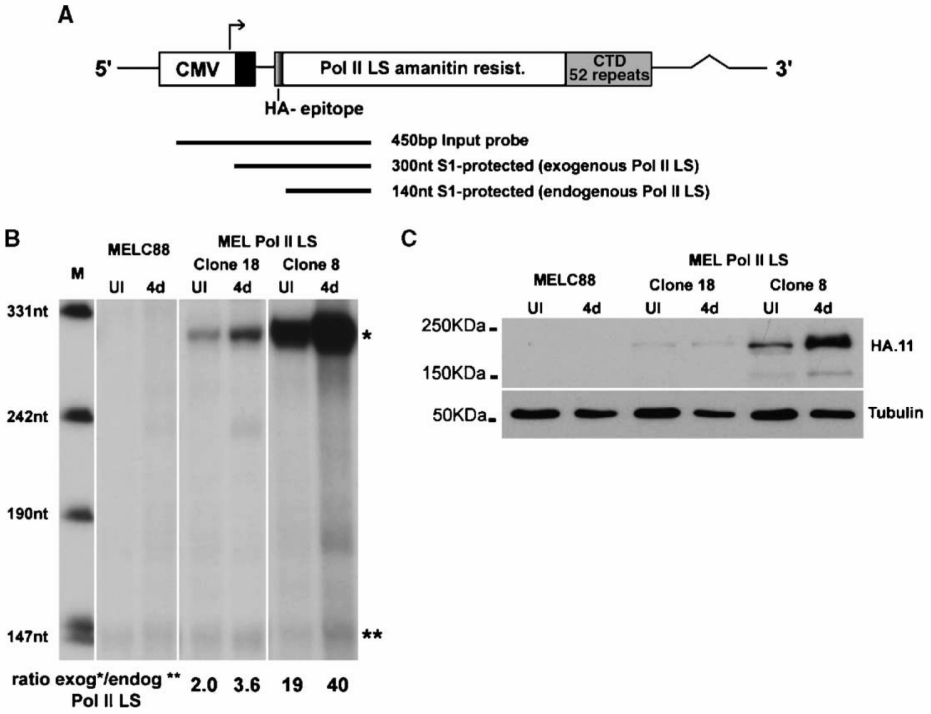


Fig. 1 – Generation of stable transfected MEL cells expressing RNA Pol II LS. (A) Schematic representation of the vector used to express α-amanitin-resistant RNA Pol II LS and probes used for S1-nuclease protection assays. Expression of RNA Pol II LS containing an HA-epitope tag is driven by a CMV promoter. The extent of the S1-nuclease protection assay probe is indicated along with the predicted protected products for the exogenous and endogenous Pol II LS. (B) Expression analysis of the endogenous and exogenous RNA Pol II LS by S1-nuclease protection assay. Total RNA fractions (40 μg) were prepared from non-transfected MEL C88 and from transfected clones 18 and 8. RNA was extracted from both undifferentiated cells (UI) and cells that were induced to differentiate for 4 days with DMSO (4d). The intensity of the bands corresponding to the exogenous (*) and endogenous (**) RNA Pol II LS were quantified by PhosphorImager and the ratio is given under each lane. (C) Expression analysis of the exogenous RNA Pol II LS by western blotting. Total protein extracts were prepared from non-transfected MEL C88 and from transfected clones 18 and 8, both before (UI) and after 4 days of induced erythroid differentiation (4d). The blot was incubated with antibodies directed against the HA-epitope present in exogenous RNA Pol II LS (HA.11) and tubulin, as a loading control.

the following secondary antibodies were used: peroxidase-conjugated goat anti-mouse IgM (1/10,000; Jackson ImmunoResearch Labs, Inc.) and peroxidase-conjugated goat anti-mouse IgG (1/3000; BioRad). Bands were visualized using ECL (Amersham Pharmacia Biotech).

Immunofluorescence

MEL cells were allowed to adhere onto poly-L-lysine-coated coverslips before fixation. Because the subcellular distribution of RNA Pol II was previously shown to be dependent upon fixation conditions [25], we used the following alternative protocols: (a) fixation with 4% paraformaldehyde, 0.1% Triton X-100, 250 mM HEPES (pH 7.4) for 10 min, and then 8% paraformaldehyde, 250 mM HEPES (pH 7.4) for 20 min. This protocol was described to ensure the best preservation of nuclear architecture [25,26]. (b) Fixation with methanol at –20°C for 10 min preceded by rinsing the coverslips in 1% paraformaldehyde in PBS [25]. This protocol was used to maximize preservation of soluble proteins in the cytoplasm. (c) Permeabilization with 0.5% Triton X-100 in CSK buffer [100 mM NaCl, 300 mM sucrose, 10 mM PIPES, 3 mM MgCl, 1 mM EGTA, pH 6.8] containing 0.1 mM PMSF for 1 min on ice, and fixation with 3.7% paraformaldehyde in CSK buffer, for 10 min at room temperature [27]. This protocol results in extraction of soluble proteins from both the cytoplasm and the nucleus.

After fixation, the cells were rinsed 4 × 5 min in PBS. Paraformaldehyde-fixed cells were permeabilized for 10 min with 0.5% Triton X-100 in PBS and rinsed 3 × 5 min in PBS. Cells were blocked for 20 min with PBS+ (PBS with 1% BSA, 0.2% fish skin gelatin; pH 7.4) and incubated with the following primary antibodies: HA.11 (1/200), ARNA-3 (1/300), 8WG16 (1/100), H5 (1/250), CC-3 (1/300) and α-hnRNPC (1/2000) for 2 h in PBS+. Detection was with FITC-conjugated goat anti-mouse IgG (1/100, Sigma) or FITC-conjugated goat anti-mouse IgM (1/200,

Jackson ImmunoResearch Labs, Inc.) diluted in PBS+. Following incubations with antibodies the cells were rinsed 4 × 5 min in PBS/0.05% Tween 20. For DNA staining, we incubated the cells for 15 min with Topro iodide (0.3 μM, Molecular Probes) in PBS. Finally, the cells were rinsed in PBS and mounted in VectaShield (Vector Laboratories).

Microscopy

Images were acquired on a Zeiss LSM 510 confocal microscope using the PlanApochromat 63×/1.4 objective. FITC fluorescence was detected using the 488-nm line of the argon ion laser, and the 543-nm line of the helium-neon laser was used to excite Topro. Images (Tagged Image files) were transferred to Adobe Photoshop for measurements of fluorescence intensity.

Results

Establishment of cell clones expressing α-amanitin-resistant RNA Pol II LS

A potent inhibitor of RNA polymerase II is the cyclic peptide α-amanitin. This toxin binds to specific residues in RNA Pol II LS preventing translocation of the enzyme along the DNA during RNA chain elongation [28]. Substitution of a single amino acid in RNA Pol II LS is sufficient to render the enzyme resistant to α-amanitin [29]. Taking advantage of this mutation, we have generated murine erythroleukemia (MEL) cell clones that express the α-amanitin-resistant form of RNA Pol II LS under the control of the human β-globin micro-locus control region (βLCR) [14]. Because co-transfected constructs can co-integrate at the same site in the genome, in some clones the βLCR and Pol II LS plasmids will integrate in a tandem fashion next to each other, which places the βLCR side-by-side to the cytomegalovirus (CMV) promoter of the Pol II LS plasmid. As the βLCR can activate heterologous genes such as the murine *Thy1* gene and the Herpes thymidine kinase promoter linked to a $G418^R$ gene (see [14,30,31]), we reasoned that its placement next to the CMV promoter would confer erythroid-specific induction. Expression of the Pol II LS transgene could thus be induced by dimethylsulfoxide (DMSO), which triggers erythroid differentiation of the MEL cells [12]. Stable clones were selected with puromycin and screened by S1-nuclease protection assay for expression of exogenous RNA Pol II LS after addition of DMSO to the culture medium for 4 days. Out of nine clones, three did not show any expression of the transgene suggesting that only the βLCR-puromycin resistance gene plasmid had integrated in these cells whereas the remaining six expressed exogenous RNA Pol II LS at different levels (Fig. 1). Clones 18 and 8, which express 3.6- and 40-fold excess levels of exogenous RNA Pol II LS, respectively (Fig. 1B), were chosen for further analysis. In both clones, the expression level of exogenous RNA Pol II LS increased after induction of erythroid differentiation (Fig. 1B, compare UI and 4d) confirming that transcription of the transgene is controlled by the βLCR. In contrast, the expression levels of the endogenous polymerase remained constant (Fig. 1B, lower band). Western blotting analysis with an antibody that recognizes the hemagglutinin (HA) epitope located in the N-terminus of the exogenous RNA Pol II LS (see Fig. 1A) shows the amount of protein present in each of the two clones that transcribe RNA from the transgene (compare Fig. 1B with Fig. 1C).

To determine if the exogenous RNA Pol II LS was functional, we analyzed the transcriptional activity of cells from clones 18 and 8 in the presence of α-amanitin. Transcriptional inhibition by α-amanitin is closely followed by the degradation of RNA Pol II LS [32]. Accordingly, when non-transfected MEL cells (C88) were treated with 2.5 μg/ml of α-amanitin for 17 h, western blotting analysis showed that the endogenous RNA Pol II LS was no longer detected whereas the exogenous α-amanitin-resistant form in clones 18 and 8 remained unaffected (Fig. 2A). As shown in Fig. 2B, the two clones expressing α-amanitin-resistant RNA Pol II LS were able to transcribe the endogenous murine β^{maj}-globin gene in the presence of α-amanitin. Furthermore, expression of the α-amanitin-resistant RNA Pol II LS was sufficient to prevent the re-localization of splicing snRNPs into enlarged speckles (Fig. 2C) that characteristically occurs in cells under conditions of reduced transcriptional activity brought about by either treatment with transcription inhibitors (reviewed in [33]) or terminal differentiation [34]. We therefore conclude that in the presence of α-amanitin, transcriptional activity in clones 18 and 8 is maintained by exogenous RNA Pol II LS.

Excess RNA Pol II LS is hypophosphorylated and accumulates in the cytoplasm as soluble protein

RNA polymerase II is hypophosphorylated when recruited to a promoter (designated RNA Pol IIA), becoming hyperphosphorylated during transcription elongation (RNA Pol IIO). Phosphorylation of the CTD results in a decrease in the electrophoretic mobility of RNA Pol II LS in SDS–PAGE. In mammals, the phosphorylated subunit, designated IIo, has an apparent *Mr* of 240,000 whereas the unphosphorylated subunit, designated IIa, has an apparent *Mr* of 214,000 (reviewed in [35]).

We performed western blotting analysis of RNA Pol II LS expressed in MEL cell clones 18 and 8 using a monoclonal antibody (mAb ARNA-3) that recognizes an epitope localized in the non-CTD part of RNA Pol II LS [17]. Although this antibody can react with both the phosphorylated (IIo) and unphosphorylated (IIa) subunits, we detected a single predominant band of ~214 kDa that corresponds to the unphosphorylated form (Fig. 3A). The results depicted in Fig. 3A show that clone 8 expresses significantly more protein than clone 18 and that both clones over-express RNA Pol II LS relative to non-transfected cells. A similar result is obtained with mAb 8WG16, which preferentially recognizes the unphosphorylated CTD [18,36–38]. This antibody detects a major band of ~214 kDa corresponding to the IIa form and multiple additional bands of higher molecular mass that may correspond to RNA Pol II LS isoforms with CTDs containing different numbers of unphosphorylated repeats [11,36,39]. In contrast, the mAb H5, which specifically recognizes phosphoserine in positions 2 and/or 5 of the CTD repeats, detects a single band of approximately 240 kDa that corresponds to the IIo form [36,37]. No difference was detected in the intensity of the IIo bands between non-transfected MEL C88 and the two cell clones over-expressing RNA Pol II LS (Fig. 3A). Similar results

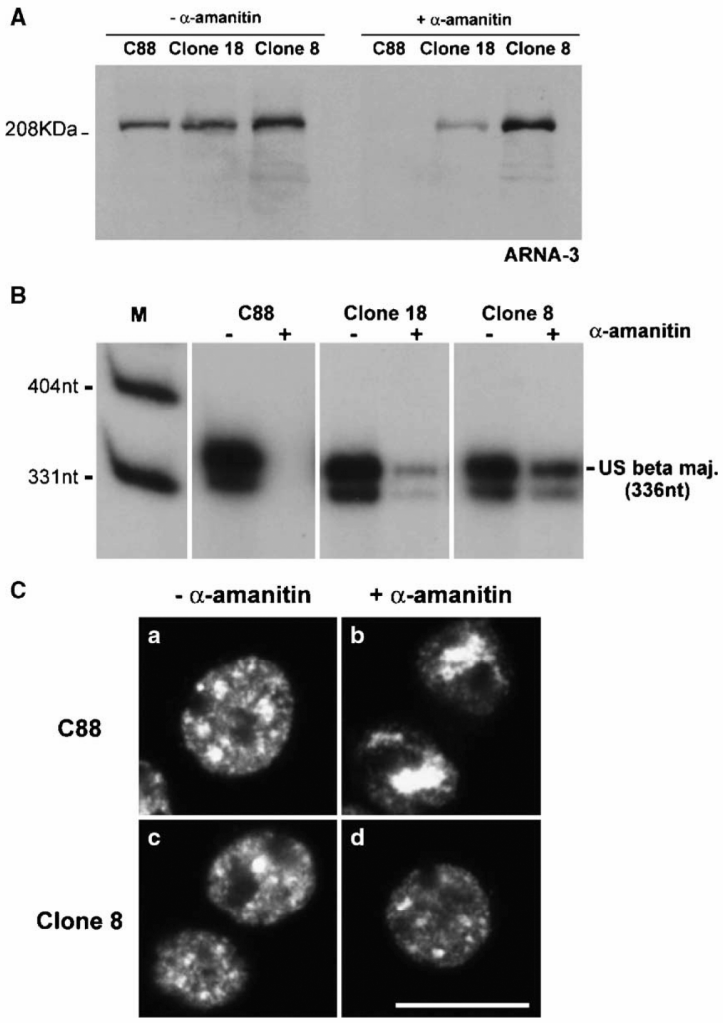


Fig. 2 – Functional analysis of the exogenous RNA Pol II LS. (A) Effect of α-amanitin on endogenous and exogenous RNA Pol II LS. Non-transfected MEL C88 cells and cells from clones 18 and 8 were induced with DMSO for 4 days and either mock treated (–) or treated with α-amanitin (2.5 μg/ml) for 17 h (+). Total cell extracts were analyzed by western blotting with mAb ARNA-3. (B) Effect of α-amanitin on transcriptional activity. Non-transfected MEL C88 cells and cells from clones 18 and 8 were induced with DMSO for 4 days and either mock treated (–) or treated with α-amanitin (2.5 μg/ml) for 17 h (+). The expression of the murine β^{maj}-globin gene was analyzed by an S1-nuclease protection assay using 10 μg of nuclear RNA with a probe that gives a protected fragment of 336nt for the pre-mRNA indicated as US (unspliced) β^{maj}. (C) Effect of α-amanitin on the distribution of splicing snRNPs in the nucleus. MEL C88 and cells from clone 8 were induced with DMSO for 4 days and either mock treated (–) or treated with α-amanitin (2.5 μg/ml) for 17 h (+). The cells were then processed for immunofluorescence with mAb Y12 that reacts with the Sm antigen present in spliceosomal snRNPs. The Y12 antibody stains the nucleoplasm in a diffuse manner with additional concentration in nuclear speckles of irregular size and shape (panels a and c). Treatment of C88 cells with α-amanitin induces a re-distribution of Sm proteins that become predominantly detected in enlarged speckles (panel b). However, in cells from clone 8 the Sm staining pattern is not affected by α-amanitin treatment (panel d). Scale bar, 10 μm.

were obtained with mAb CC-3 that also recognizes phosphoserine in position 2 of the CTD [22,36].

Next, we analyzed the subcellular distribution of RNA Pol II LS by immunofluorescence (Fig. 3B). In non-transfected C88 MEL cells, the mAb ARNA-3 labelled predominantly the nucleoplasm with a punctuate pattern (Fig. 3B, panel a). In clones 18 and 8, the nuclear pattern was similar, but additional cytoplasmic labelling was observed (Fig. 3B, panels b and c). Similar results were observed with an anti-HA-tag antibody that specifically recognizes the exogenous protein (Fig. 3B, panels d–f). Western blotting analysis of nuclear and cytoplasmic fractions isolated from non-transfected MEL C88 and

clones 18 and 8 further confirms the accumulation of exogenous RNA Pol II LS in the cytoplasm of transfected cells (Fig. 3C).

To investigate the phosphorylation status of RNA Pol II LS forms that accumulate in the cytoplasm, we performed immunofluorescence and western blotting experiments with monoclonal antibodies mAb 8WG16 (recognizes the IIa form) and mAb H5 (detects the IIo form). The immunofluorescence staining produced by mAb 8WG16 in non-transfected C88 cells is predominantly nuclear, whereas in clones 18 and 8 there is clear cytoplasmic labelling (Fig. 4A). The intensity of the cytoplasmic signal is significantly higher in clone 8 than in clone 18 (Fig. 4A). In contrast, the staining pattern produced by mAb H5 is exclusively nuclear in C88, clone 18 and clone 8 (Fig. 4B). Moreover, the staining intensity is similar in the three types of cells (Fig. 4B). Western blotting analysis using mAb 8WG16 confirms that clones 18 and 8 have increased amounts of form IIa both in the nucleus and in the cytoplasm (Fig. 4C), whereas mAb H5 detects similar levels of the IIo form present exclusively in the nuclear fractions (Fig. 4D).

It has been previously shown that the unphosphorylated IIa form of RNA Pol II is more easily extracted than the hyperphosphorylated form (IIo) [20,40]. This prompted us to perform immunofluorescence with mAb 8WG16 in cells that were either immediately fixed or treated with Triton X-100 before formaldehyde fixation. As shown in Fig. 4E, treatment with Triton X-100 prior to fixation completely abolished staining in the cytoplasm and reduced labelling in the nucleus.

Taken together these results indicate that excess RNA Pol II LS produced in clones 18 and 8 fails to be hyperphosphorylated at Ser^2 and accumulates both in the nucleus and in the cytoplasm as soluble protein.

RNA Pol II LS is exported to the cytoplasm by a CRM1-dependent pathway

The observed accumulation of excess RNA Pol II LS in the cytoplasm of clones 18 and 8 could be due to saturation of the transport mechanism responsible for import into the nucleus of the protein newly synthesized in the cytoplasm. Alternatively, the RNA Pol II LS molecules produced in clones 18 and 8 could be imported into the nucleus but the excess would then be exported back to the cytoplasm. The later hypothesis implies that RNA Pol II LS is a shuttling protein. To explore this possibility, we treated erythroid-induced cells from the high exogenous Pol II LS expressing clone 8 with leptomycin B (LMB), a drug that specifically blocks the nuclear export activity of CRM1 [41,42]. The cells were immunolabelled with

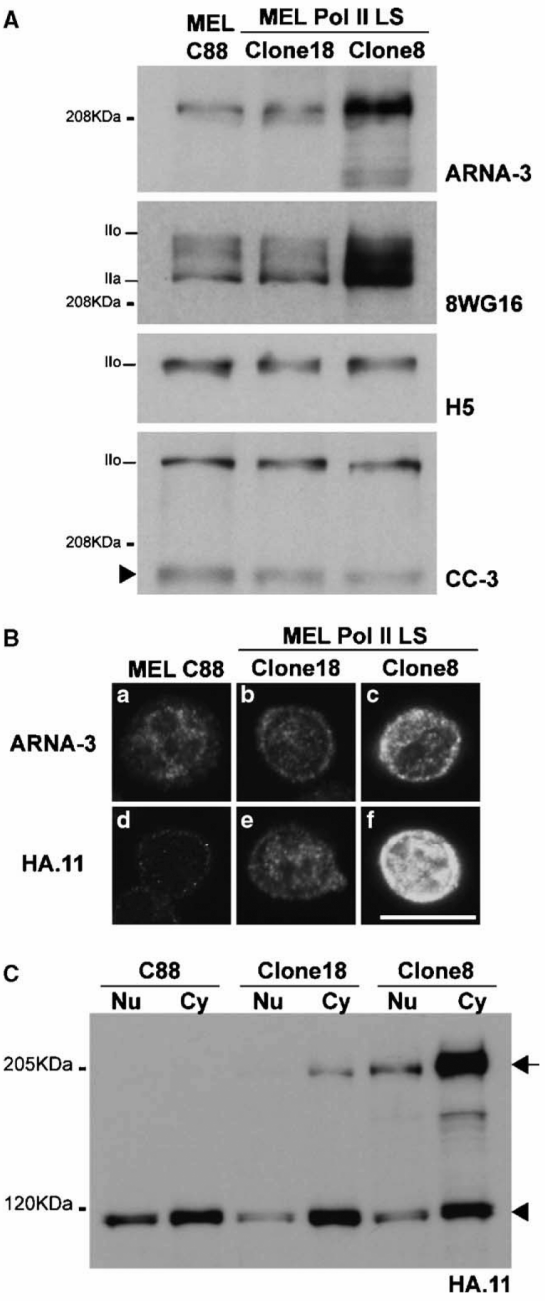


Fig. 3 – Exogenous RNA Pol II LS is differentially phosphorylated. (A) Western blotting analysis of total cell extracts from non-transfected MEL C88 and clones 18 and 8 induced with DMSO for 4 days. The blots were incubated with mAbs ARNA-3, 8WG16, H5 and CC-3, as indicated. Note that mAb CC-3 reveals an additional band (arrowhead) that was described to correspond to cross-reacting phosphoproteins present in cellular extracts [48, 49]. (B) Immunofluorescence analysis of non-transfected MEL C88 and clones 18 and 8 induced with DMSO for 4 days. The cells were rinsed in 1% paraformaldehyde in PBS, fixed with methanol at –20°C for 10 min and incubated with mAbs ARNA-3 (panels a–c) or HA.11 (panels d–f). Note a peripheral rim staining corresponding to the cytoplasm in panels b, c, e and f. Scale bar, 10 μm. (C) Western blotting analysis of nuclear and cytoplasmic fractions isolated from non-transfected MEL C88 and clones 18 and 8 induced with DMSO for 4 days. The blot was incubated with mAb HA.11, which reacts with the exogenous RNA Pol II LS (arrow) and an additional unspecific band that is also present in non-transfected MEL C88 cells (arrowhead).

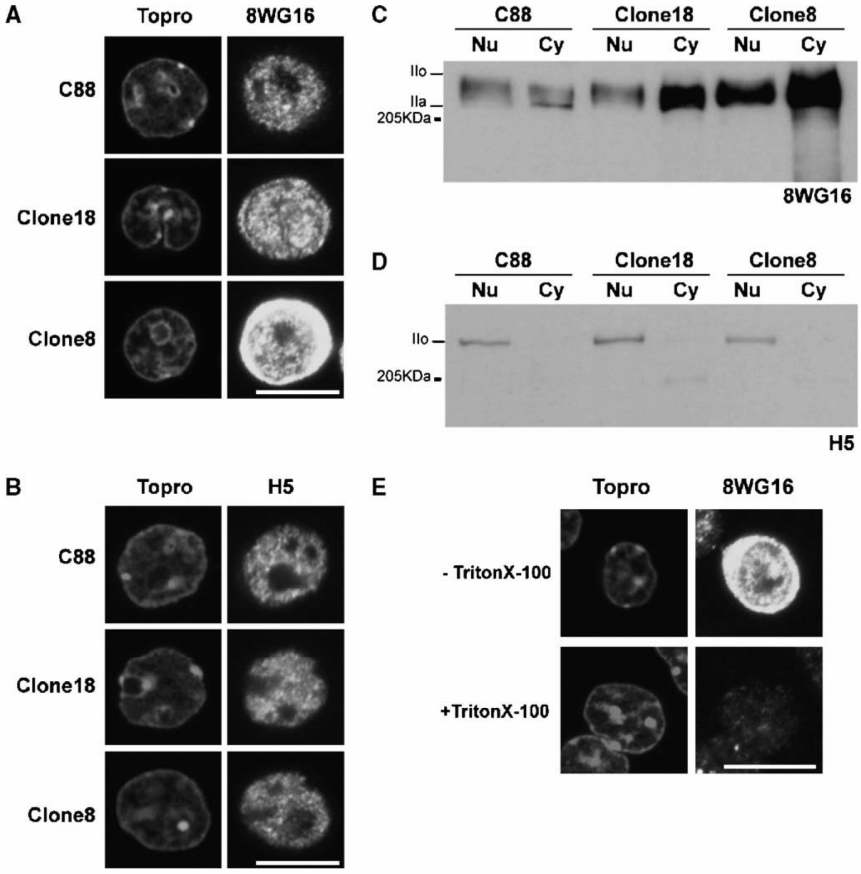


Fig. 4 – Subcellular distribution of IIa and IIo forms of RNA Pol II LS. Non-transfected MEL C88 and clones 18 and 8 were induced with DMSO for 4 days. (A) The cells were rinsed in 1% paraformaldehyde in PBS, fixed with methanol at –20°C for 10 min and double-labelled with the DNA stain Topro to visualize the nucleus and mAb 8WG16. Note the peripheral rim staining corresponding to unphosphorylated forms of RNA Pol II LS in the cytoplasm of transfected cells. (B) The cells were fixed with 4% formaldehyde, 0.1% Triton X-100, 250 mM HEPES (pH 7.4) for 10 min, and then in 8% formaldehyde, 250 mM HEPES (pH 7.4) for 20 min and permeabilized for 10 min with 0.5% Triton X-100 in PBS. The cells were double-labelled with Topro and mAb H5. Note the absence of phosphorylated polymerase in the cytoplasm of transfected cells. (C and D) Western blotting analysis of nuclear and cytoplasmic fractions with mAb 8WG16 (C) and mAb H5 (D). (E) The cells were either rinsed in 1% paraformaldehyde in PBS and fixed with methanol at –20°C for 10 min (–Triton X-100) or incubated in 0.5% Triton X-100 in CSK buffer 1 min on ice and fixed with 3.7% paraformaldehyde in CSK buffer for 10 min at room temperature (+Triton X-100). The cells were double-labelled with Topro and mAb 8WG16. Bars, 10 μm.

mAb 8WG16 and fluorescence intensity in the cytoplasm was quantified. The proportion of cells with cytoplasmic fluorescence above an arbitrary threshold was estimated and plotted for each treatment (Fig. 5A). The results show that LMB caused a progressive reduction in the proportion of cells accumulating RNA Pol II in the cytoplasm, suggesting that CRM1 is required for export of the protein to the cytoplasm. This prompted us to analyze the effect of LMB treatment on endogenous RNA Pol II. Non-transfected MEL C88 cells were erythroid-induced for 4 days and then treated with LMB for 3 h and immunolabelled with mAb 8WG16. As a control, the cells were stained with a mAb that recognizes hnRNP C, a protein that does not shuttle to the cytoplasm [43]. After 3 h of treatment with LBM, the nuclear fluorescence intensity produced by mAb 8WG16 was significantly higher than in untreated cells, whereas the nuclear fluorescence intensity produced by anti-hnRNP C was not affected (Fig. 5B). This suggests that although at steady state RNA Pol II LS is exclusively detected in the nucleus, a fraction of the protein is constantly shuttling back and forth between the nucleus and the cytoplasm. However, when uninduced MEL cells were treated with LMB for 3–5 h, no significant increase in the nuclear fluorescence intensity produced by mAb 8WG16 was detected and similar results were obtained in HeLa cells (data not shown). A possible explanation for this discrepancy relies on the well-known phenomenon that erythroid differentiation reduces transcriptional activity of many genes (see [34]). This will create a surplus of RNA Pol II LS in the nucleus that

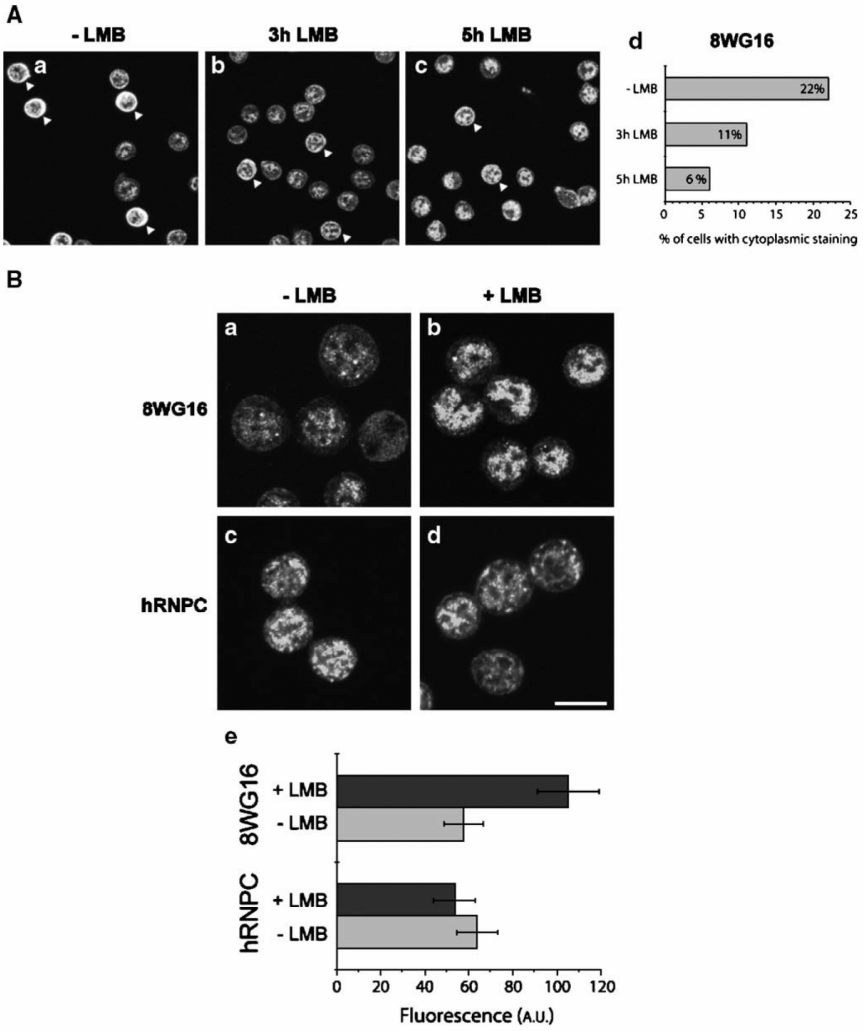


Fig. 5 – CRM1-mediated export of RNA Pol II LS. (A) Cells from clone 8 were induced to differentiate for 4 days and either mock treated (–LMB) or treated with LMB (50 nM) for 3 and 5 h. The cells were rinsed in 1% formaldehyde in PBS, fixed with methanol at –20°C for 10 min and immunolabelled with mAb 8WG16 (panels a–c). Two separate experiments were performed for each treatment, with a total of 300–400 cells analyzed per experiment. Cells with cytoplasmic fluorescence intensity above an arbitrary threshold are indicated (arrowheads). The proportion of these cells in each population is plotted in panel d. (B) Non-transfected MEL C88 cells were induced to differentiate for 4 days and either mock treated (–LMB) or treated with 50 nM LMB for 3 h (+LMB). The cells were fixed as described in A and immunolabelled with mAbs 8WG16 (panels a, b) and α-hnRNPC (panels c, d). Three independent experiments were performed, with a total of 20–25 cells analyzed per experiment. Means ± SE are plotted in panel e. The nuclear fluorescence is expressed in arbitrary units (A.U.). Scale bar, 10 μm.

shuttles to the cytoplasm. In contrast, in actively transcribing cells, most molecules of RNA Pol II LS are engaged in transcriptional complexes and therefore are not exported to the cytoplasm.

Discussion

In this work, we have established murine erythroleukemia (MEL) cell clones that express an α-amanitin-resistant form of RNA polymerase II largest subunit under the control of the human β-globin micro-locus control region. In order to obtain maximal levels of transgene expression, we analyzed cells at 4 days after induction of erythroid differentiation. Consistent with previous reports [32], treatment of the parental non-transfected cells with α-amanitin induces degradation of the RNA Pol II LS protein and completely abolishes transcriptional activity by RNA polymerase II. In contrast, in transfected clones treated with α-amanitin the protein encoded by the transgene is not degraded and

supports transcription of endogenous β^{maj}-globin pre-mRNA. Although cells from transfected clone 8 express up to 40-fold excess RNA Pol II LS relative to the endogenous level (Fig. 1B), the amount of β^{maj}-globin pre-mRNA produced in these cells is clearly reduced after addition of α-amanitin (Fig. 2B). This indicates that exogenous RNA Pol II LS is able to functionally replace the endogenous protein in vivo albeit with a lower efficacy.

Several lines of evidence indicate that transcriptional inhibitors such as α-amanitin, actinomycin D or DRB induce a re-organization of splicing factors in the nucleus with concentration in enlarged speckles (for a recent review, see [33]). Speckles are dynamic subnuclear domains from which pre-messenger RNA splicing factors cycle continuously to sites of active transcription. Upon inactivation of transcription, the association of splicing factors with nuclear speckles is accentuated. This effect is illustrated in Fig. 2C, which shows the distribution of splicing snRNPs in non-transfected C88 MEL cells before (panel a) and after (panel b) treatment with α-amanitin for 17 h. In contrast, the drug fails to affect the subnuclear distribution of snRNPs in cells expressing the α-amanitin-resistant form of RNA Pol II LS (Fig. 2C, panels c and d), confirming that re-localization of splicing factors is caused by lack of transcriptional activity and not by an indirect effect of α-amanitin treatment.

Immunofluorescence analysis using either ARNA-3 antibody (that recognizes both the endogenous and exogenous RNA Pol II LS) or anti-HA antibody (to detect exclusively the exogenous protein) shows punctuate staining diffusely distributed throughout the nucleoplasm (Fig. 3B). This is in agreement with previous reports using antibodies that label both hypo- and hyperphosphorylated forms of RNA Pol II LS [25,44,45]. Although some studies found RNA polymerase II co-localizing with splicing factors in nuclear speckles [20,44,46], more recent work demonstrated that staining of these subnuclear domains by anti-RNA polymerase II antibodies can be induced by fixation protocols [25].

The IIa and IIo phosphoisoforms of RNA Pol II LS coexist in a dynamic equilibrium, depending on the antagonistic action of specific CTD kinases and phosphatases [35]. Here we observe that over-expression of RNA Pol II LS results in accumulation of the IIa isoform. Unlike the IIo form, the amount of which remained constant and bound in the nucleus, the over-expressed unphosphorylated form of RNA Pol II LS accumulated both in the nucleus and in the cytoplasm and was mostly released when cells were extracted with Triton X-100 prior to fixation. Thus, it appears that the number of RNA Pol II LS molecules available to assemble into active polymerases is tightly regulated in the cell nucleus.

The presence of over-expressed RNA Pol II in the cytoplasm was previously reported in cells transfected with an EGFP fusion version of RNA Pol II LS [47]. The accumulation of over-expressed RNA Pol II LS in the cytoplasm could be due to inefficient import into the nucleus of the excess protein synthesized in the cytoplasm. Alternatively, RNA Pol II LS molecules could enter the nucleus but those that are in excess would be exported back into the cytoplasm. To investigate whether RNA Pol II LS molecules can shuttle between the nucleus and the cytoplasm, we compared the subcellular distribution of both endogenous and over-expressed RNA Pol II LS before and after treatment of erythroid-induced MEL cells with leptomycin B (LMB), a specific inhibitor of CRM1-dependent nuclear export [41,42]. Our results show that LMB reduces the cytoplasmic accumulation of over-expressed RNA Pol II LS and increases the nuclear concentration of the endogenous protein in differentiated MEL cells. This strongly suggests that RNA Pol II LS can shuttle between the nucleus and the cytoplasm. Possibly, nucleo-cytoplasmic shuttling of RNA polymerase II largest subunit represents a mechanism to regulate the pool of available molecules in the nucleus and thereby ensure correct assembly of transcriptionally active enzyme complexes.

Acknowledgments

We are grateful to O. Bensaude (Ecole Normale Superieure, Paris) for critical reading of the manuscript and to José Braga and José Rino for help with some experiments. We also thank the following groups for generously providing reagents used in this work: W. Schaffner (University of Zürich) for the CTD construct; M. Vincent (Université Laval) for the CC-3 antibody; G. Dreyfuss (Howard Hughes Medical Institute) for monoclonal antibody 4F4; and A. Lamond (University of Dundee) for Y12 monoclonal antibody. This work was supported by Fundação para a Ciência e Tecnologia, Portugal, and by the European Science Foundation (EuroDYNA Programme).

REFERENCES

[1] A. Shilatifard, R.C. Conaway, J.W. Conaway, The RNA polymerase II elongation complex, Annu. Rev. Biochem. 72 (2003) 693–715.
[2] J.L. Corden, D.L. Cadena, J.M. Ahearn Jr., M.E. Dahmus, A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II, Proc. Natl. Acad. Sci. U. S. A. 82 (1985) 7934–7938.
[3] J.L. Corden, Tails of RNA polymerase II, Trends Biochem. Sci. 15 (1990) 383–387.
[4] M.L. West, J.L. Corden, Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations, Genetics 140 (1995) 1223–1233.
[5] O. Bensaude, F. Bonnet, C. Casse, M.F. Dubois, V.T. Nguyen, B. Palancade, Regulated phosphorylation of the RNA polymerase II C-terminal domain (CTD), Biochem. Cell. Biol. 77 (1999) 249–255.
[6] W.Y. Kim, M.E. Dahmus, Immunochemical analysis of mammalian RNA polymerase II subspecies. Stability and relative in vivo concentration, J. Biol. Chem. 261 (1986) 14219–14225.
[7] H. Lu, O. Flores, R. Weinmann, D. Reinberg, The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex, Proc. Natl. Acad. Sci. U. S. A. 88 (1991) 10004–10008.
[8] J.D. Chesnut, J.H. Stephens, M.E. Dahmus, The interaction of RNA polymerase II with the adenovirus-2 major late promoter is precluded by phosphorylation of the C-terminal domain of subunit IIa, J. Biol. Chem. 267 (1992) 10500–10506.
[9] P. Komarnitsky, E.J. Cho, S. Buratowski, Different phosphorylated forms of RNA polymerase II and associated

mRNA processing factors during transcription, Genes Dev. 14 (2000) 2452–2460.
[10] H. Cho, T.K. Kim, H. Mancebo, W.S. Lane, O. Flores, D. Reinberg, A protein phosphatase functions to recycle RNA polymerase II, Genes Dev. 13 (1999) 1540–1552.
[11] M.S. Kobor, J. Archambault, W. Lester, F.C. Holstege, O. Gileadi, D.B. Jansma, E.G. Jennings, F. Kouyoumdjian, A.R. Davidson, R.A. Young, J. Greenblatt, An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in *S. cerevisiae*, Mol. Cell 4 (1999) 55–62.
[12] A. Antoniou, Induction of erythroid-specific expression in murine erythroleukemia (MEL) cells, in: E.J. Murray (Ed.), Methods in Molecular Biology: Gene transfer and Expression Protocols, vol. 7, The Human Press Inc., Clinfton, NJ, 1991, pp. 421–434.
[13] H.P. Gerber, M. Hagmann, K. Seipel, O. Georgiev, M.A. West, Y. Litingtung, W. Schaffner, J.L. Corden, RNA polymerase II C-terminal domain required for enhancer-driven transcription, Nature 374 (1995) 660–662.
[14] P. Collis, M. Antoniou, F. Grosveld, Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression, EMBO J. 9 (1990) 233–240.
[15] M. Antoniou, F. Geraghty, J. Hurst, F. Grosveld, Efficient 3′-end formation of human β-globin mRNA in vivo requires sequences within the last intron but occurs independently of the splicing reaction, Nucleic Acids Res. 26 (1998) 721–729.
[16] E. de Boer, M. Antoniou, V. Mignotte, L. Wall, F. Grosveld, The human beta-globin promoter; nuclear protein factors and erythroid specific induction of transcription, EMBO J. 7 (1988) 4203–4212.
[17] A. Kramer, R. Haars, R. Kabisch, H. Will, F.A. Bautz, E.K. Bautz, Monoclonal antibody directed against RNA polymerase II of *Drosophila melanogaster*, Mol. Gen. Genet. 180 (1980) 193–199.
[18] N.E. Thompson, T.H. Steinberg, D.B. Aronson, R.R. Burgess, Inhibition of in vivo and in vitro transcription by monoclonal antibodies prepared against wheat germ RNA polymerase II that react with the heptapeptide repeat of eukaryotic RNA polymerase II, J. Biol. Chem. 264 (1989) 11511–11520.
[19] S.L. Warren, A.S. Landolfi, C. Curtis, J.S. Morrow, Cytostellin: a novel, highly conserved protein that undergoes continuous redistribution during the cell cycle, J. Cell Sci. 103 (Pt 2) (1992) 381–388.
[20] D.B. Bregman, L. Du, S. vanderZee, S.L. Warren, Transcription-dependent redistribution of the large subunit of RNA polymerase II to discrete nuclear domains, J. Cell Biol. 129 (1995) 287–298.
[21] A. Thibodeau, M. Vincent, Monoclonal antibody CC-3 recognizes phosphoproteins in interphase and mitotic cells, Exp. Cell Res. 195 (1991) 145–153.
[22] M. Vincent, P. Lauriault, M.F. Dubois, S. Lavoie, O. Bensaude, B. Chabot, The nuclear matrix protein p255 is a highly phosphorylated form of RNA polymerase II largest subunit which associates with spliceosomes, Nucleic Acids Res. 24 (1996) 4649–4652.
[23] E.A. Lerner, M.R. Lerner, C.A. Janeway Jr., J.A. Steitz, Monoclonal antibodies to nucleic acid-containing cellular constituents: probes for molecular biology and autoimmune disease, Proc. Natl. Acad. Sci. U. S. A. 78 (1981) 2737–2741.
[24] Y.D. Choi, G. Dreyfuss, Monoclonal antibody characterization of the C proteins of heterogeneous nuclear ribonucleoprotein complexes in vertebrate cells, J. Cell Biol. 99 (1984) 204–1997.
[25] P.V. Guillot, S.Q. Xie, M. Hollinshead, A. Pombo, Fixation-induced redistribution of hyperphosphorylated RNA polymerase II in the nucleus of human cells, Exp. Cell Res. 295 (2004) 460–468.
[26] A. Pombo, P. Cuello, W. Schul, J.B. Yoon, R.G. Roeder, P.R. Cook, S. Murphy, Regional and temporal specialization in the nucleus: a transcriptionally-active nuclear domain rich in PTF, Oct1 and PIKA antigens associates with specific chromosomes early in the cell cycle, EMBO J. 17 (1998) 1768–1778.
[27] E.G. Fey, G. Krochmalnic, S. Penman, The nonchromatin substructures of the nucleus: the ribonucleoprotein (RNP)-containing and RNP-depleted matrices analyzed by sequential fractionation and resinless section electron microscopy, J. Cell Biol. 102 (1986) 1654–1665.
[28] D.A. Bushnell, P. Cramer, R.D. Kornberg, Structural basis of transcription: alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 1218–1222.
[29] M.S. Bartolomei, J.L. Corden, Localization of an alpha-amanitin resistance mutation in the gene encoding the largest subunit of mouse RNA polymerase II, Mol. Cell. Biol. 7 (1987) 586–594.
[30] G. BlomvanAssendelft, O. Hanscombe, F. Grosveld, D.R. Greaves, The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner, Cell 56 (1989) 969–977.
[31] M. Antoniou, F. Grosveld, beta-globin dominant control region interacts differently with distal and proximal promoter elements, Genes Dev. 4 (1990) 1007–1013.
[32] V.T. Nguyen, F. Giannoni, M.F. Dubois, S.J. Seo, M. Vigneron, C. Kedinger, O. Bensaude, In vivo degradation of RNA polymerase II largest subunit triggered by alpha-amanitin, Nucleic Acids Res. 24 (1996) 2924–2929.
[33] A.I. Lamond, D.L. Spector, Nuclear speckles: a model for nuclear organelles, Nat. Rev., Mol. Cell Biol. 4 (2003) 605–612.
[34] M. Antoniou, M. Carmo-Fonseca, J. Ferreira, A.I. Lamond, Nuclear organization of splicing snRNPs during differentiation of murine erythroleukemia cells in vitro, J. Cell Biol. 123 (1993) 1055–1068.
[35] B. Palancade, O. Bensaude, Investigating RNA polymerase II *carboxyl*-terminal domain (CTD) phosphorylation, Eur. J. Biochem. 270 (2003) 3859–3870.
[36] M. Patturajan, R.J. Schulte, B.M. Sefton, R. Berezney, M. Vincent, O. Bensaude, S.L. Warren, J.L. Corden, Growth-related changes in phosphorylation of yeast RNA polymerase II, J. Biol. Chem. 273 (1998) 4689–4694.
[37] J.C. Jones, H.P. Phatnani, T.A. Haystead, J.A. MacDonald, S.M. Alam, A.L. Greenleaf, C-terminal repeat domain kinase I phosphorylates Ser2 and Ser5 of RNA polymerase II C-terminal domain repeats, J. Biol. Chem. 279 (2004) 24957–24964.
[38] E.J. Cho, T. Takagi, C.R. Moore, S. Buratowski, mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain, Genes Dev. 11 (1997) 3319–3326.
[39] J.G. Valay, M. Simon, M.F. Dubois, O. Bensaude, C. Facca, G. Faye, The KIN28 gene is required both for RNA polymerase II mediated transcription and phosphorylation of the Rpb1p CTD, J. Mol. Biol. 249 (1995) 535–544.
[40] F. Bonnet, M. Vigneron, O. Bensaude, M.F. Dubois, Transcription-independent phosphorylation of the RNA polymerase II C-terminal domain (CTD) involves ERK kinases (MEK1/2), Nucleic Acids Res. 27 (1999) 4399–4404.
[41] M. Fornerod, M. Ohno, M. Yoshida, I.W. Mattaj, CRM1 is an export receptor for leucine-rich nuclear export signals, Cell 90 (1997) 1051–1060.
[42] N. Kudo, N. Matsumori, H. Taoka, D. Fujiwara, E.P. Schreiner, B. Wolff, M. Yoshida, S. Horinouchi, Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 9112–9117.
[43] S. Pinol-Roma, G. Dreyfuss, Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm, Nature 355 (1992) 730–732.

[44] C. Zeng, E. Kim, S.L. Warren, S.M. Berget, Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity, EMBO J. 16 (1997) 1401–1412.
[45] M.A. Grande, I. van der Kraan, L. de Jong, R. van Driel, Nuclear distribution of transcription factors in relation to sites of transcription and RNA polymerase II, J. Cell Sci. 110 (Pt 15) (1997) 1781–1791.
[46] D.B. Bregman, L. Du, Y. Li, S. Ribisi, S.L. Warren, Cytostellin distributes to nuclear regions enriched with splicing factors, J. Cell Sci. 107 (Pt 3) (1994) 387–396.
[47] K. Sugaya, M. Vigneron, P.R. Cook, Mammalian cell lines expressing functional RNA polymerase II tagged with the green fluorescent protein, J. Cell Sci. 113 (Pt 15) (2000) 2679–2683.
[48] S.B. Lavoie, A.L. Albert, H. Handa, M. Vincent, O. Bensaude, The peptidyl-prolyl isomerase Pin1 interacts with hSpt5 phosphorylated by Cdk9, J. Mol. Biol. 312 (2001) 675–685.
[49] A.L. Albert, S.B. Lavoie, M. Vincent, Multisite phosphorylation of Pin1-associated mitotic phosphoproteins revealed by monoclonal antibodies MPM-2 and CC-3, BMC Cell Biol. 5 (2004) 22.

CHAPTER III

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

Mechanisms must exist in the nucleus to proofread the transcripts and to ensure that only fully processed transcripts reach the translation machinery in the cytoplasm. The main goal of this work was to elucidate the quality control mechanism responsible for the nuclear retention of human β-globin transcripts with mutations that impaired their processing. The system we have exploited in these studies consists of wild-type and processing mutant versions of the human β-globin gene stably transfected into murine erythroleukemia (MEL) cells, under the control of the locus control region (LCR). The principal discovery of this study was the demonstration that pre-mRNA mutants unable to be exported to the cytoplasm due to the inability to undergo splicing or 3' end processing are retained within the nucleus in close proximity to the site of transcription. This observation imply that efficient pre-mRNA processing is crucial and therefore rate limiting for the release of transcripts from the site of transcription. We showed that exon junction complex (EJC) proteins co-localise with protein components of the spliceosome at the transcription sites in the nucleoplasm, providing *in vivo* evidence that the EJC binds co-transcriptionally to mRNPs. In contrast, mutant β-globin pre-mRNAs that are neither spliced nor released from the site of transcription fail to recruit spliceosome snRNPs and EJC proteins. This suggests that splicing of β-globin pre-mRNA is required for efficient recruitment of the EJC and subsequent targeting of the resulting mRNP to the nuclear export pathway. A major player in co-transcriptional pre-mRNA maturation is the carboxyl-terminal domain (CTD) of the largest subunit of RNA Polymerase II (RNA Pol II LS) which acts by facilitating specific interactions between processing factors while the transcript is still attached to the polymerase (reviewed in Bentley, 2005). We generated MEL cell lines that express α-amanitin resistant RNA Pol II LS with either full-length or truncated forms of the CTD to investigate the role of the CTD in transcript release. The results revealed that deleting 21 C-terminal heptads of the CTD causes human β-globin mRNA retention at the site of transcription but without inhibiting splicing or 3' end formation. This implies a novel and previously unsuspected involvement of the CTD in mRNP maturation events that occur after splicing, cleavage and polyadenylation have taken place.

Transcription site retention of pre-mRNA processing mutants

The finding that both splicing and polyadenylation factors can associate with CTD of RNA Pol II (Du and Warren, 1997; McCracken et al., 1997b; Mortillaro et al., 1996; Yuryev et al., 1996) let to the hypothesis that RNA processing mutants may be retained close to the DNA template by remaining tethered to a stalled or abnormal processing machinery

associated with the CTD of the polymerase. The range of pre-mRNA mutants we have analysed suggests that assembly of the processing machinery for both splicing and 3' end formation is involved in the retention mechanism. A prediction of the model in which both the splicing and the 3' end processing machinery is associated with the CTD of RNA Pol II is that, a normal mRNA would be released from the CTD as it is processed, whereas a pre-mRNA processing mutant would remain bound to the processing machinery and therefore to the CTD (Figure 10). As a consequence RNA Pol II would not be released from the mutant gene upon transcription termination, which would result in the stalling of oncoming RNA Pol II molecules on the template and therefore in the reduction of the overall rate of transcription.

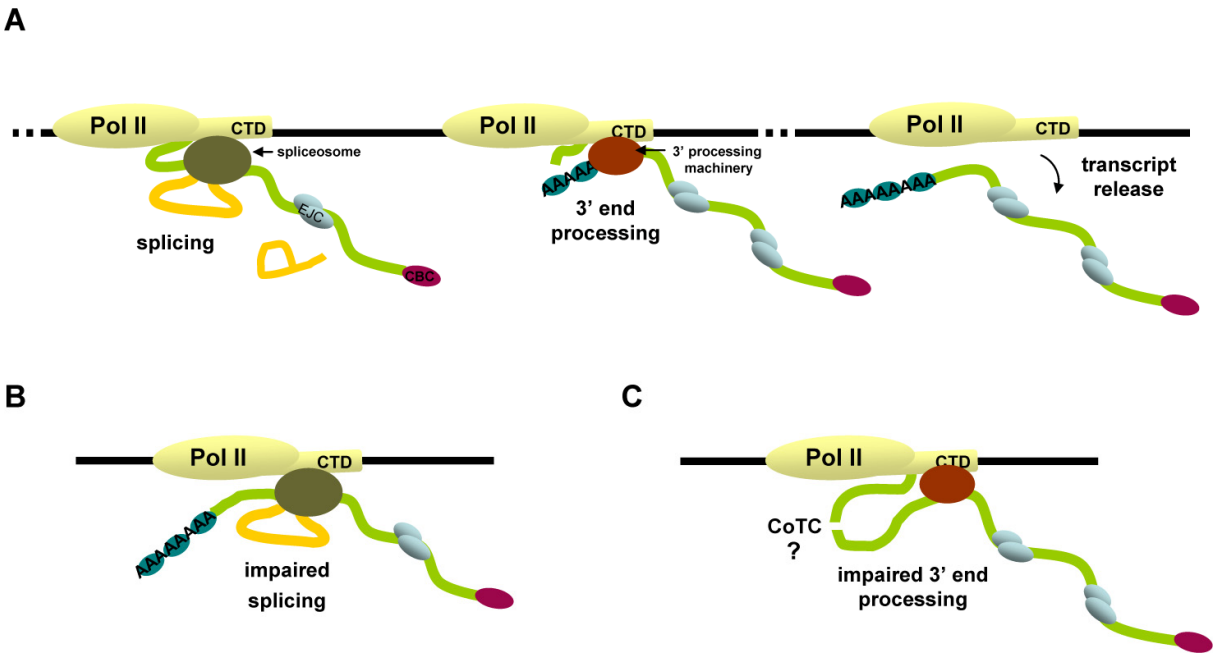


Figure 10 - A model for processing-dependent mRNA release from the transcription site **(A)** Both splicing and 3' end processing factors associate with the CTD of RNA Pol II and a fully processed mRNA is released from the CTD after completion of splicing and 3' end processing. Human β-globin transcripts harbouring mutations that impair splicing **(B)** or 3' end processing **(C)** are retained near the site of transcription, presumably stalled by the processing machinery tethered to the CTD.

Furthermore, if pre-mRNA processing mutant transcripts fail to be released from the site of transcription they should accumulate to a significantly higher amount at the site of transcription when compared to the transcripts of the wild-type human β-globin gene and this should result in an increase in the intensity of the nuclear foci produced by *in situ* hybridisation. However, a careful examination of the wild-type MELβWT and 5' splice site

mutant MELβSM cells that harbour a similar transgene copy number showed that there is no increase in signal intensity. In fact, nuclear foci of the βSM transcripts are frequently smaller and less intense than the βWT foci. This result is consistent with the hypothesis that mutated transcripts are stalled on the template with a concurrent feedback mechanism to the transcription machinery.

To test this hypothesis, the effect of this splicing mutation on transcription should be carefully analysed. This could be achieved comparing the density of transcribing RNA Pol II along the βWT and βSM transcription units using a nuclear run-on assay. Chromatin immunoprecipitation (ChIP) assays with antibodies that recognise the different phosphorylation forms of the CTD could also be used to compare the evolution of CTD phosphorylation along the βWT and βSM transcription units. This experiment could give valuable information into the dynamics of the human β-globin transcription cycle because it is well established that the CTD is dynamically phosphorylated during each transcription cycle. Five out of the seven residues of the CTD heptapeptides can be phosphorylated, but residues Ser^2 and Ser^5 are the main targets of phosphorylation (Bensaude et al., 1999; West and Corden, 1995). When RNA Pol II is recruited to a promoter the CTD is hypophosphorylated (recognised by 8WG16 antibody) and becomes hyperphosphorylated during transcription elongation (Kim and Dahmus, 1986). Phosphorylation of Ser^5 residues (recognised by H14 antibody) occurs at or near the promoter, while Ser^2 residues (recognised by H5 antibody) are phosphorylated during the transition to the elongation phase of transcription as the polymerases move away from the promoter (Chesnut et al., 1992; Komarnitsky et al., 2000; Lu et al., 1991). CTD phosphorylated on Ser^2 residues is considered a marker of efficiently transcribing or elongating RNA Pol II (Palancade and Bensaude, 2003). Therefore comparison of ChIP results obtained with H5 antibody on MELβWT and mutant MELβSM cells could indicate if the splicing mutation has an effect on transcription, specifically in the transition from the initiation to the elongation phase.

The site of transcription: a hot spot for quality control

Current evidence suggests that any failure in pre-mRNA processing compromising the integrity of an mRNA may cause its retention in the nucleus at the site of transcription (reviewed in Jensen et al., 2003). Why is such a surveillance mechanism operating at the site of transcription? There is evidence that pre-mRNA processing is a co-transcriptional event and that the different steps in pre-mRNA processing are all interconnected (reviewed by

Bentley, 1999). Most likely, this cross-talk between the distinct processing steps during mRNA maturation plays a critical role to guarantee the quality of the final messenger, ensuring that only correctly processed transcripts leave the site of transcription.

Efficient 5' capping enhances both splicing of the first intron and 3' end processing, while 3' end cleavage requires splicing of the terminal intron and splicing of the last intron depends on the recognition of the poly(A) site (see Fong and Zhou, 2001). The presence of an intact poly(A) signal is also crucial for transcription termination, i.e. release of the RNA transcript from the polymerase and dissociation of the polymerase from the DNA template (Osheim et al., 1999). This constitutes an additional quality checkpoint ensuring that transcription does not terminate until RNA Pol II has moved past the end of the entire mRNA sequence. In fact, transcription termination has been shown to occur hundreds of kilobases past the poly(A) site (Bauren et al., 1998; Dye and Proudfoot, 1999; Osheim et al., 1999). Cleavage at the poly(A) site occurs at the same time as or just before transcription termination (Dye and Proudfoot, 1999). Termination additionally requires the presence of a transcriptional pause site in the DNA (Yonaha and Proudfoot, 2000). This pause may be necessary for the interaction of the CTD-associated processing factors with the pre-mRNA, which may then trigger the cleavage reaction at the 3' end and promote the release of the polymerase from the DNA template (Dye and Proudfoot, 1999).

Analysis of the human β-globin gene indicated that transcription termination is preceded by multiple cleavage reactions in the co-transcriptional cleavage (CoTC) site downstream of the poly(A) site. The occurrence of such pre-termination cleavage reactions, appears to be essential for effective release of the nascent transcript from the DNA template (Dye and Proudfoot, 2001). The RNA sequence at the CoTC forms a ribozyme and mutations that inhibit self cleavage at the CoTC site also inhibit transcriptional termination (Teixeira et al., 2004). According to these data, the nascent transcript probably maintains an indirect interaction with the CTD during co-transcriptional cleavage, until the final cleavage and polyadenylation reaction triggers both the release of the transcript from the CTD and the dissociation of RNA Pol II from the template

The time lag between transcription of the poly(A) site and the final cleavage reaction that precedes polyadenylation may have important functional implications. In fact, by allowing the completion of all the processing reactions before termination, this delay may be crucial for the final check on the integrity of the transcript before cleavage occurs. Consistent with this view, a mutation in the 3' splice site of the terminal intron of human β-globin

abolishes transcriptional termination (Dye and Proudfoot, 1999). Since efficient splicing of the last intron of β-globin is essential for proper cleavage (Antoniou et al., 1998; Dye and Proudfoot, 1999), the effect of this splicing mutation on transcript termination may be a consequence of its interference with 3' end cleavage. Nevertheless, transcription termination appears to constitute the final trigger for the release of a fully processed mRNA from the site of transcription and failure of any of the pre-mRNA processing reactions may impair this final step, therefore retaining the transcripts at the gene locus. The β-globin 5' splice site mutant βSM we have analysed is 3' end cleaved at normal rates (Antoniou et al., 1998). It remains to be determined if transcription termination may also be impaired in this mutant and if this is contributing for its retention at the site of transcription. We have also shown that the CTD is important for transcription site release independently of its role in promoting splicing and 3' end processing. It would also be important to determine if the CTD has a role in transcription termination. Interestingly, it was shown in *S. cerevisiae* that Pcf11, a component of the cleavage factor IA (CFIA), associates directly with both RNA and the CTD and is able to dismantle the elongating complex *in vitro* (Zhang et al., 2005). In *Drosophila* this protein (dPcf11) is also directly involved in termination (Zhang and Gilmour, 2006). dPcf11 is concentrated at the 3' end of the hsp70 gene, and depletion of dPcf11 with RNAi causes RNA Pol II to readthrough the normal region of termination. Biochemical analysis revealed that dPcf11 forms a bridge between the CTD and the RNA and dismantles the elongation complexes by a CTD-dependent mechanism (Zhang and Gilmour, 2006). Human Pcf11 (hPcf11) is a component of cleavage factor II (CFII), which is required for pre-mRNA cleavage during 3' end processing (de Vries et al., 2000). However, functions of hPcf11 in RNA Pol II termination have only recently started to be explored. It was shown that it causes premature termination on an HIV provirus (Zhang et al., 2007) and depletion of hPcf11 in HeLa cells reduces the efficiency of transcriptional termination on transfected constructs and stabilizes the 3' products of RNA cleavage at or beyond the poly(A) signal (West and Proudfoot, 2007).

In the yeast *S. cerevisiae*, several types of defective mRNAs also accumulate at or near the site of transcription (reviewed in Saguez et al., 2005). In all cases of transcription site retention in yeast the Rrp6 protein as well as other components of the nuclear exosome are required, suggesting that this complex is part of a quality control checkpoint that monitors for correct processing of pre-mRNA (Dunn et al., 2005; Hilleren et al., 2001; Jensen et al., 2001b; Libri et al., 2002). The exosome is a complex of 3' to 5' exoribonucleases composed of ten

core subunits and the Rrp6 protein which is required for its nuclear activity (reviewed in Vanacova and Stefl, 2007). There is evidence in *Drosophila* that the exosome is recruited to active genes via interactions with elongation factors (Andrulis et al., 2002). The human counterpart of the yeast exosome is the polymyositis-scleroderma (PM/Scl) complex and the protein PM/Scl-100 the orthologue of Rrp6p. However, a role for PM/Scl-100 in transcription-site retention has not been described. We have shown that the protein PM/Scl-100 is recruited *in vivo* to the site of human β-globin transcription, both when it is transcribed by an RNA Pol II LS with wild-type CTD and when transcription occurs exclusively by an RNA Pol II LS with 31 repeats in the CTD which results in transcription-site retention. It will be important to determine if in mammals transcription site retention of processing defective mutants also occurs in the absence of a functional nuclear exosome.

Formation of an export competent mRNP particle

During processing, nascent mRNA molecules assemble together with RNA binding proteins into ribonucleoprotein particles (mRNPs; Aguilera, 2005; Moore, 2005). Once released from the gene template, mRNP particles must reach the nuclear pore in order to be translocated to the cytoplasm. It is currently believed that after disengaging from the transcription and processing machinery mRNPs move by random Brownian motion throughout the nucleoplasm (reviewed by Gorski et al., 2006). As diffusion cannot be regulated, traffic control of mRNP molecules must rely on retention at specific locations within the nucleus (Gorski et al., 2006). According to the current view these locations are the site of transcription (reviewed in Jensen et al., 2003) and, at least in yeast, the nuclear pore (Galy et al., 2004).

The commitment of a transcript for export from the nucleus involves the binding of several RNA binding proteins that include shuttling hnRNPs, EJC proteins and export factors. Translocation of mRNPs through the nuclear pore complex requires binding of an heterodimer composed of NXF1 (or TAP) and NXT1 (or p15) (Mex67 and Mtr2 in *S.cerevisiae*). NXF1, which interacts both with RNA binding adapter proteins and components of the nuclear pore complex, is believed to be the major receptor for the export of mRNA molecules to the cytoplasm (Izaurralde, 2002; Lei and Silver, 2002b; Reed and Hurt, 2002). The EJC protein REF1/Aly has been identified as a factor that facilitates the nuclear export of mRNA by interacting with the export receptor NXF1 (Rodrigues et al., 2001; Stutz et al., 2000; Zhou et al., 2000). REF1/Aly is recruited to spliced mRNAs through physical

interactions with UAP56 (Luo et al., 2001; Strasser and Hurt, 2001), a putative RNA helicase also implicated as a splicing factor required for early spliceosome assembly (Libri et al., 2001; Zhang and Green, 2001). Contrasting with NXF1 and UAP56, the depletion of which causes nuclear accumulation of mRNA, the EJC proteins REF/Aly, RNPS1, SRm160, and Y14 are dispensable for bulk mRNA export (Gatfield and Izaurralde, 2002; Longman et al., 2003; MacMorris et al., 2003). Accordingly, SR proteins have been proposed as adaptors that in addition to REF1/Aly can mediate the interaction between NXF1 and cellular mRNAs (Huang et al., 2003; Huang and Steitz, 2001; Huang et al., 2004). Interestingly, SR proteins have recently been shown to specifically associate with RNA Pol II and function in coupling transcription to splicing (Das et al., 2007).

Since the formation of an export-competent mRNP particle involves the binding of several proteins to the nascent transcripts and the protein composition changes as the transcripts are being processed it is possible that these proteins are a key component in the quality control mechanism that prevents the abnormal transcripts from reaching the cytoplasm. These proteins could prevent the release of incorrectly processed transcripts from specific locations and/or promote the export of the correct ones by the addition of an "export licence" that may be important for the translocation step at the nuclear pore complex. In this context one hypothesis to explain the retention human β-globin processing mutants at the transcription site could be the absence of recruitment of proteins essential for the release and/or transport of the transcripts from the gene locus. We studied the recruitment of EJC proteins and mRNA export factors to human β-globin transcription sites and found that several EJC proteins co-localise with protein components of the spliceosome at sites of transcription and splicing in the nucleoplasm, providing *in vivo* evidence that the EJC binds co-transcriptionally to mRNPs. No accumulation of the export factors NXF1 and NXT1 was detected at the transcription sites, arguing that in mammalian cells these proteins are recruited at a later stage in the export pathway probably shortly before or after release from the gene template. In agreement with this result it was reported that NXF1 binds very weakly to purified spliced mRNPs *in vitro* (Zhou et al., 2000). Contrasting with the results obtained for wild-type human β-globin transcripts, mutant pre-mRNAs defective in splicing and 3' end processing do not recruit spliceosome snRNPs or the EJC proteins SRm160, REF and UAP56. Although these results suggest that splicing of β-globin pre-mRNAs is required for efficient recruitment of the EJC and subsequent targeting of the resulting mRNP to the nuclear export pathway further studies are required to determine if these proteins are involved

in transcription site release. One possibility to address this would be to artificially target EJC proteins and/or export factors to the retained processing mutants and ask which protein could induce the release of the transcripts.

CTD dependent mRNA release from the site of transcription

The CTD of RNA Pol II LS is composed of heptapeptide repeats with the consensus YSPTSPS (Corden et al., 1985), and the number of repeats varies from 26 in yeast to 40 in *Drosophila* and 52 in mammals (Allison et al., 1985; Corden et al., 1985). Not only has the number of repeats increased through evolution but also the number of repeats that deviates from the consensus sequence (Stiller and Hall, 2002). In mammals there are 31 non-consensus repeats and most of them are located at the C-terminus. After repeat 52 the mammalian CTD further comprises a specific 10 amino acid motif. It has been speculated that the evolution of longer CTDs with non-consensus repeats could have resulted in the attribution of specific functions to different regions of the CTD (Fong and Bentley, 2001). Accordingly, CTD deletion analysis has shown that heptad repeats 1-15 or 1-25 support capping but not splicing or 3' end formation, whereas heptads 27-52 plus the C-terminal 10 residues can support efficient capping, splicing and 3' end formation (Fong and Bentley, 2001). More recent studies have demonstrated that mutation of the terminal 10 amino acid motif of the CTD inhibited splicing, 3' end cleavage (Fong et al., 2003) and RNA release from the site of transcription (Bird et al., 2005). Based on these observations it was proposed that the CTD is required for transcript release as a consequence of its role in splicing and 3' end cleavage (Bird et al., 2005). However, we have show that a partial truncation of the CTD (Δ31) containing heptads 1-23, 36-38 and 48-52 including the terminal 10 amino acid motif is sufficient to support transcription, splicing, 3' end cleavage and polyadenylation, but the mRNA fails to be efficiently released from the site of transcription. These novel observations imply that the CTD is involved in processes that control the release of transcripts by a mechanism independent from splicing and cleavage.

An important player acting at the interface between transcription and biogenesis of export-competent mRNPs is the TREX complex. This complex, initially characterised in yeast is formed by the THO (a four-protein complex involved in transcription elongation and the maintenance of genome stability) (Chávez et al., 2000), Sub2p and Yra1p (Strasser et al., 2002). Null mutations of any component of THO lead to impairment of transcription and RNA export defects (Chávez et al., 2000; Rondón et al., 2003; Schneiter et al., 1999; Strasser

et al., 2002). Mutations of Sub2 and Yra1 confer similar transcription-impairment phenotypes as THO mutations (Fan et al., 2001; Jimeno et al., 2002; Rondón et al., 2003). Recently, the human TREX complex was characterised in detail and shown to contain the EJC proteins REF1/Aly, UAP56 and the human counterpart of the THO complex (Masuda et al., 2005). *In vitro* studies indicate that the human THO complex associates with spliced mRNA but not with unspliced pre-mRNA and this association is independent of transcription (Masuda et al., 2005). Whether the TREX complex participates in release of mRNA from the site of transcription is unknown. We showed that the EJC proteins REF1/Aly and SRm160 are recruited to human β-globin transcripts made by the RNA Pol II with the Δ31 CTD, that remain at the transcription site after transcriptional inhibition, indicating that the deposition of the EJC onto spliced mRNA is not sufficient for its release.

Several lines of evidence argue that 3' end formation by cleavage and polyadenylation is important for mRNA export. Nevertheless, artificial 3' end cleavage by a self-cleaving ribozyme is not sufficient to permit mRNA export (Eckner et al., 1991; Huang and Carmichael, 1996; Libri et al., 2002). Moreover, accumulation of ribozyme-cleaved RNA near its site of transcription was rescued by a synthetic A-tail (a stretch of DNA-encoded adenosine residues) in the absence of factors required for normal cleavage and polyadenylation, suggesting that mRNA release may be mediated by proteins that bind to poly(A)$^+$ RNA such as Pab1p and Nab2p (Dower et al., 2004). Indeed, Nab2p has been implicated in 3' end formation and yeast strains defective for Nab2p function accumulate hyperadenylated RNA in the nucleus (Hector et al., 2002) similarly to what is observed in mutants for general mRNA export factors (Hilleren et al., 2001; Jensen et al., 2001b). Although no direct and prominent role for yeast Pab1p or mammalian PABPN1 in mRNA export has yet been demonstrated, Pab1p overexpression rescued the export defect of the Nab2p-defective strain further suggesting that Pab1p may promote export or remove retention factors (Dower et al., 2004). Thus, one possibility is that the CTD plays a role in recruiting polyadenylation factors and poly(A)$^+$ RNA binding proteins, which in turn contribute for mRNA clearance from the site of transcription. Nevertheless, we showed that the retained transcripts made by RNA Pol II LS with the Δ31 CTD are correctly polyadenylated.

Although the mechanism via which the CTD is involved in release of the mRNA from the transcription site is unknown, we propose that the missing heptads in the truncated Δ31 CTD mutant are required for binding of proteins implicated in a final co-transcriptional maturation of spliced and 3' end cleaved mRNAs into export-competent ribonucleoprotein

particles (Figure 11). The next challenge will be to determine the proteins that are differentially recruited by the wild-type and the truncated CTD in order to determine which protein or protein complex is responsible for that final co-transcriptional maturation step that allows the release of the transcripts from the gene template. One good candidate to fulfil this role is Pcf11, a protein that is part of the 3' end processing machinery and has recently emerged as a key factor in the transcription termination processs (Sadowski et al., 2003; West and Proudfoot, 2007; Zhang et al., 2005; Zhang and Gilmour, 2006). This protein was shown to associates directly with both RNA and the CTD in yeast (Zhang et al., 2005) and *Drosophila* (Zhang and Gilmour, 2006). Is able to promote the dismantle the elongation complexes *in vitro* by a CTD-dependent mechanism and is also required for transcription termination *in vivo* (Sadowski et al., 2003; West and Proudfoot, 2007; Zhang et al., 2005; Zhang and Gilmour, 2006).

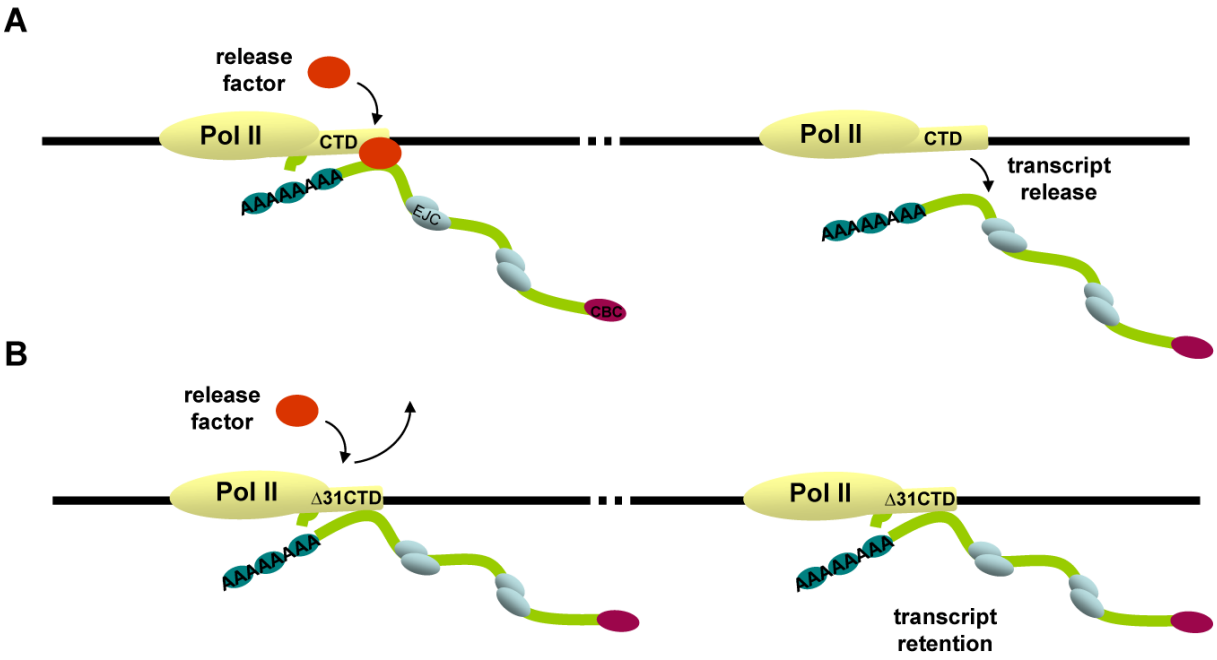


Figure 11 - A model for CTD-dependent mRNA release from the transcription site (A) Fully processed transcripts are released from the DNA template possibly upon recruitment of proteins (release factors) implicated in a final co-transcriptional maturation step of spliced and 3' end cleaved mRNAs into export-competent ribonucleoprotein particles. **(B)** The missing heptads in the truncated Δ31 CTD mutant are required for binding of proteins implicated in the release of fully processed transcripts. When an mRNA is made by an RNA Pol II with a Δ31 CTD these release factors are not recruited and the mRNA is retained at the transcription site.

Final considerations

mRNA quality control mechanisms safeguard cells from the accumulation of malfunctioning proteins therefore modulating the clinical manifestations of many genetic disorders. Although a number of mRNA quality control mechanisms have been identified in recent years, the molecular players behind these mechanisms are still largely uncharacterised. In particular, it is generally unclear how these quality control systems can distinguish between correct and defective mRNAs and mRNP complexes. Uncovering the molecular mechanisms of the different mRNA surveillance pathways and understanding how these systems are interconnected will be a challenge for the future. Because mRNA quality control mechanisms modulate the clinical outcome of many genetic disorders uncovering their molecular players may represent promising targets for future therapeutic intervention.

REFERENCES

Abruzzi, K.C., Lacadie, S. and Rosbash, M. (2004) Biochemical analysis of TREX complex recruitment to intronless and intron-containing yeast genes. *EMBO J*, **23**, 2620-2631.

Aguilera, A. (2005) Cotranscriptional mRNP assembly: from the DNA to the nuclear pore. *Curr Opin Cell Biol*, **17**, 242-250.

Allen, T.D., Cronshaw, J.M., Bagley, S., Kiseleva, E. and Goldberg, M.W. (2000) The nuclear pore complex: mediator of translocation between nucleus and cytoplasm. *J Cell Sci*, **113 (Pt 10)**, 1651-1659.

Allison, L.A., Moyle, M., Shales, M. and Ingles, C.J. (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell*, **42**, 599-610.

Allison, L.A., Wong, J.K., Fitzpatrick, V.D., Moyle, M. and Ingles, C.J. (1988) The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a conserved structure with an essential function. *Mol Cell Biol*, **8**, 321-329.

Andrecka, J., Lewis, R., Bruckner, F., Lehmann, E., Cramer, P. and Michaelis, J. (2008) Single-molecule tracking of mRNA exiting from RNA polymerase II. *Proc Natl Acad Sci U S A*, **105**, 135-140.

Andrulis, E.D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P. and Lis, J.T. (2002) The RNA processing exosome is linked to elongating RNA polymerase II in Drosophila. *Nature*, **420**, 837-841.

Ansari, A. and Hampsey, M. (2005) A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. *Genes Dev*, **19**, 2969-2978.

Antonioiu, A. (1995) Clinical defects in pre-mRNA processing. In Angus, A.I. (ed.), *Pre-mRNA Processing*. R.G. Landes Company, pp. 187-201.

Antoniou, A. (1991) Induction of erythroid-specific expression in murine erythroleukemia (MEL) cells. In Murray, E.J. (ed.), *Methods in Molecular Biology : Gene transfer and Expression Protocols*. The Human Press Inc., Clinfton, NJ, Vol. 7, pp. 421-434.

Antoniou, M., Geraghty, F., Hurst, J. and Grosveld, F. (1998) Efficient 3'-end formation of human ß-globin mRNA in vivo requires sequences within the last intron but occurs independently of the splicing reaction. *Nucleic Acids Res*, **26**, 721-729.

Armache, K.J., Kettenberger, H. and Cramer, P. (2003) Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc Natl Acad Sci U S A*, **100**, 6964-6968.

Ashe, H.L., Monks, J., Wijgerde, M., Fraser, P. and Proudfoot, N.J. (1997) Intergenic transcription and transinduction of the human beta-globin locus. *Genes Dev*, **11**, 2494-2509.

Ast, G. (2004) How did alternative splicing evolve? *Nat Rev Genet*, **5**, 773-782.

Barron-Casella, E. and Corden, J.L. (1992) Conservation of the mammalian RNA polymerase II largest-subunit C- terminal domain. *J Mol Evol*, **35**, 405-410.

Bass, B.L. (2002) RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem*, **71**, 817-846.

Bauren, G., Belikov, S. and Wieslander, L. (1998) Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3'-end formation and excision of the 3'-terminal intron. *Genes Dev*, **12**, 2759-2769.

Bensaude, O., Bonnet, F., Casse, C., Dubois, M.F., Nguyen, V.T. and Palancade, B. (1999) Regulated phosphorylation of the RNA polymerase II C-terminal domain (CTD). *Biochem Cell Biol*, **77**, 249-255.

Bentley, D. (1999) Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr Opin Cell Biol*, **11**, 347-351.

Bentley, D.L. (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol*, **17**, 251-256.

Bird, G., Fong, N., Gatlin, J.C., Farabaugh, S. and Bentley, D.L. (2005) Ribozyme cleavage reveals connections between mRNA release from the site of transcription and pre-mRNA processing. *Mol Cell*, **20**, 747-758.

Bird, G., Zorio, D.A. and Bentley, D.L. (2004) RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol*, **24**, 8963-8969.

Birse, C.E., Minvielle-Sebastia, L., Lee, B.A., Keller, W. and Proudfoot, N.J. (1998) Coupling termination of transcription to messenger RNA maturation in yeast. *Science*, **280**, 298-301.

Blackwood, E.M. and Kadonaga, J.T. (1998) Going the distance: a current view of enhancer action. *Science*, **281**, 60-63.

Blencowe, B.J. (2006) Alternative splicing: new insights from global analyses. *Cell*, **126**, 37-47.

Blobel, G. (1985) Gene gating: a hypothesis. *Proc Natl Acad Sci U S A*, **82**, 8527-8529.

Blom van Assendelft, G., Hanscombe, O., Grosveld, F. and Greaves, D.R. (1989) The beta-globin dominant control region activates homologous and heterologous promoters in a tissue-specific manner. *Cell*, **56**, 969-977.

Boeger, H., Bushnell, D.A., Davis, R., Griesenbeck, J., Lorch, Y., Strattan, J.S., Westover, K.D. and Kornberg, R.D. (2005) Structural basis of eukaryotic gene transcription. *FEBS Lett*, **579**, 899-903.

Bousquet-Antonelli, C., Presutti, C. and Tollervey, D. (2000) Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell*, **102**, 765-775.

Braun, I.C., Herold, A., Rode, M., Conti, E. and Izaurralde, E. (2001) Overexpression of TAP/p15 heterodimers bypasses nuclear retention and stimulates nuclear mRNA export. *J Biol Chem*, **276**, 20536-20543.

Bray, M., Prasad, S., Dubay, J.W., Hunter, E., Jeang, K.T., Rekosh, D. and Hammarskjold, M.L. (1994) A small element from the Mason-Pfizer monkey virus genome makes human immunodeficiency virus type 1 expression and replication Rev-independent. *Proc Natl Acad Sci U S A*, **91**, 1256-1260.

Briggs, M.W., Burkard, K.T. and Butler, J.S. (1998) Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3' end formation. *J Biol Chem*, **273**, 13255-13263.

Brodsky, A.S. and Silver, P.A. (2000) Pre-mRNA processing factors are required for nuclear export. *RNA*, **6**, 1737-1749.

Bungert, J., Dave, U., Lim, K.C., Lieuw, K.H., Shavit, J.A., Liu, Q. and Engel, J.D. (1995) Synergistic regulation of human beta-globin gene switching by locus control region elements HS3 and HS4. *Genes Dev*, **9**, 3083-3096.

Buratowski, S. (2003) The CTD code. *Nat Struct Biol*, **10**, 679-680.

Buratowski, S. (2005) Connections between mRNA 3' end processing and transcription termination. *Curr Opin Cell Biol*, **17**, 257-261.

Burgess, S.M. and Guthrie, C. (1993) A mechanism to enhance mRNA splicing fidelity: the RNA-dependent ATPase Prp16 governs usage of a discard pathway for aberrant lariat intermediates. *Cell*, **73**, 1377-1391.

Burkard, K.T. and Butler, J.S. (2000) A nuclear 3'-5' exonuclease involved in mRNA degradation interacts with Poly(A) polymerase and the hnRNA protein Npl3p. *Mol Cell Biol*, **20**, 604-616.

Bushnell, D.A. and Kornberg, R.D. (2003) Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc Natl Acad Sci U S A*, **100**, 6969-6973.

Butler, J.S. (2002) The yin and yang of the exosome. *Trends Cell Biol*, **12**, 90-96.

Caceres, J.F. and Kornblihtt, A.R. (2002) Alternative splicing: multiple control mechanisms and involvement in human disease. *Trends Genet*, **18**, 186-193.

Caceres, J.F., Screaton, G.R. and Krainer, A.R. (1998) A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. *Genes Dev*, **12**, 55-66.

Calapez, A., Pereira, H.M., Calado, A., Braga, J., Rino, J., Carvalho, C., Tavanez, J.P., Wahle, E., Rosa, A.C. and Carmo-Fonseca, M. (2002) The intranuclear mobility of messenger RNA binding proteins is ATP dependent and temperature sensitive. *J Cell Biol*, **159**, 795-805.

Callebaut, I., Moshous, D., Mornon, J.P. and de Villartay, J.P. (2002) Metallo-beta-lactamase fold within nucleic acids processing enzymes: the beta-CASP family. *Nucleic Acids Res*, **30**, 3592-3601.

Cao, A. and Moi, P. (2002) Regulation of the globin genes. *Pediatr Res*, **51**, 415-421.

Carter, D., Chakalova, L., Osborne, C.S., Dai, Y.F. and Fraser, P. (2002) Long-range chromatin regulatory interactions in vivo. *Nat Genet*, **32**, 623-626.

Chabot, B., Bisotto, S. and Vincent, M. (1995) The nuclear matrix phosphoprotein p255 associates with splicing complexes as part of the [U4/U6.U5] tri-snRNP particle. *Nucleic Acids Res*, **23**, 3206-3213.

Chang, D.D. and Sharp, P.A. (1989) Regulation by HIV Rev depends upon recognition of splice sites. *Cell*, **59**, 789-795.

Chapman, R.D., Heidemann, M., Albert, T.K., Mailhammer, R., Flatley, A., Meisterernst, M., Kremmer, E. and Eick, D. (2007) Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7. *Science*, **318**, 1780-1782.

Chávez, S., Beilharz, T., Rondon, A.G., Erdjument-Bromage, H., Tempst, P., Svejstrup, J.Q., Lithgow, T. and Aguilera, A. (2000) A protein complex containing Tho2, Hpr1, Mft1 and a novel protein, Thp2, connects transcription elongation with mitotic recombination in Saccharomyces cerevisiae. *EMBO J*, **19**, 5824-5834.

Cheng, C. and Sharp, P.A. (2003) RNA polymerase II accumulation in the promoter-proximal region of the dihydrofolate reductase and gamma-actin genes. *Mol Cell Biol*, **23**, 1961-1967.

Cheng, H., Dufu, K., Lee, C.S., Hsu, J.L., Dias, A. and Reed, R. (2006) Human mRNA export machinery recruited to the 5' end of mRNA. *Cell*, **127**, 1389-1400.

Chesnut, J.D., Stephens, J.H. and Dahmus, M.E. (1992) The interaction of RNA polymerase II with the adenovirus-2 major late promoter is precluded by phosphorylation of the C-terminal domain of subunit IIa. *J Biol Chem*, **267**, 10500-10506.

Cho, E.J., Rodriguez, C.R., Takagi, T. and Buratowski, S. (1998) Allosteric interactions between capping enzyme subunits and the RNA polymerase II carboxy-terminal domain. *Genes Dev*, **12**, 3482-3487.

Cho, E.J., Takagi, T., Moore, C.R. and Buratowski, S. (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes Dev*, **11**, 3319-3326.

Cho, H., Kim, T.K., Mancebo, H., Lane, W.S., Flores, O. and Reinberg, D. (1999) A protein phosphatase functions to recycle RNA polymerase II. *Genes Dev*, **13**, 1540-1552.

Choi, O.R. and Engel, J.D. (1988) Developmental regulation of beta-globin gene switching. *Cell*, **55**, 17-26.

Collis, P., Antoniou, M. and Grosveld, F. (1990) Definition of the minimal requirements within the human beta-globin gene and the dominant control region for high level expression. *EMBO J*, **9**, 233-240.

Comer, F.I. and Hart, G.W. (2001) Reciprocity between O-GlcNAc and O-phosphate on the carboxyl terminal domain of RNA polymerase II. *Biochemistry*, **40**, 7845-7852.

Connelly, S. and Manley, J.L. (1988) A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes Dev*, **2**, 440-452.

Cook, P.R. (1999) The organization of replication and transcription. *Science*, **284**, 1790-1795.

Cooke, C. and Alwine, J.C. (2002) Characterization of specific protein-RNA complexes associated with the coupling of polyadenylation and last-intron removal. *Mol Cell Biol*, **22**, 4579-4586.

Cooke, C., Hans, H. and Alwine, J.C. (1999) Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal. *Mol Cell Biol*, **19**, 4971-4979.

Cooper, G.M. and Hausman, R.E. (2007) *The Cell: A Molecular Approach*. Fourth edition. ASM press, Washington, D.C.

Coppola, J.A., Field, A.S. and Luse, D.S. (1983) Promoter-proximal pausing by RNA polymerase II in vitro: transcripts shorter than 20 nucleotides are not capped. *Proc Natl Acad Sci U S A*, **80**, 1251-1255.

Corden, J.L., Cadena, D.L., Ahearn, J.M., Jr. and Dahmus, M.E. (1985) A unique structure at the carboxyl terminus of the largest subunit of eukaryotic RNA polymerase II. *Proc Natl Acad Sci U S A*, **82**, 7934-7938.

Cramer, P. (2004) RNA polymerase II structure: from core to functional complexes. *Curr Opin Genet Dev*, **14**, 218-226.

Cramer, P., Bushnell, D.A. and Kornberg, R.D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science*, **292**, 1863-1876.

Cramer, P., Caceres, J.F., Cazalla, D., Kadener, S., Muro, A.F., Baralle, F.E. and Kornblihtt, A.R. (1999) Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol Cell*, **4**, 251-258.

Cramer, P., Pesce, C.G., Baralle, F.E. and Kornblihtt, A.R. (1997) Functional association between promoter structure and transcript alternative splicing. *Proc Natl Acad Sci U S A*, **94**, 11456-11460.

Cremer, T. and Cremer, C. (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet*, **2**, 292-301.

Cullen, B.R. (2003) Nuclear RNA export. *J Cell Sci*, **116**, 587-597.

Custódio, N., Antoniou, M. and Carmo-Fonseca, M. (2006) Abundance of the largest subunit of RNA polymerase II in the nucleus is regulated by nucleo-cytoplasmic shuttling. *Exp Cell Res*, **312**, 2557-2567.

Custódio, N. and Carmo-Fonseca, M. (2001) Quality control of gene expression in the nucleus. *J Cell Mol Med*, **5**, 267-275.

Custódio, N., Carmo-Fonseca, M., Geraghty, F., Pereira, H.S., Grosveld, F. and Antoniou, M. (1999) Inefficient processing impairs release of RNA from the site of transcription. *EMBO J*, **18**, 2855-2866.

Custódio, N., Carvalho, C., Condado, I., Antoniou, M., Blencowe, B.J. and Carmo-Fonseca, M. (2004) In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei. *RNA*, **10**, 622-633.

Custódio, N., Vivo, M., Antoniou, M. and Carmo-Fonseca, M. (2007) Splicing- and cleavage-independent requirement of RNA polymerase II CTD for mRNA release from the transcription site. *J Cell Biol*, **179**, 199-207.

Dantonel, J.C., Murthy, K.G., Manley, J.L. and Tora, L. (1997) Transcription factor TFIID recruits factor CPSF for formation of 3' end of mRNA. *Nature*, **389**, 399-402.

Darnell, J.E., Philipson, L., Wall, R. and Adesnik, M. (1971) Polyadenylic acid sequences: role in conversion of nuclear RNA into messenger RNA. *Science*, **174**, 507-510.

Darzacq, X., Singer, R.H. and Shav-Tal, Y. (2005) Dynamics of transcription and mRNA export. *Curr Opin Cell Biol*, **17**, 332-339.

Das, B., Butler, J.S. and Sherman, F. (2003) Degradation of normal mRNA in the nucleus of Saccharomyces cerevisiae. *Mol Cell Biol*, **23**, 5502-5515.

Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P. and Reed, R. (2007) SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell*, **26**, 867-881.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D. and Kornblihtt, A.R. (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell*, **12**, 525-532.

de Vries, H., Ruegsegger, U., Hubner, W., Friedlein, A., Langen, H. and Keller, W. (2000) Human pre-mRNA cleavage factor II(m) contains homologs of yeast proteins and bridges two other cleavage factors. *EMBO J*, **19**, 5895-5904.

Degot, S., Le Hir, H., Alpy, F., Kedinger, V., Stoll, I., Wendling, C., Seraphin, B., Rio, M.C. and Tomasetto, C. (2004) Association of the breast cancer protein MLN51 with the exon junction complex via its speckle localizer and RNA binding module. *J Biol Chem*, **279**, 33702-33715.

Denis, M.M., Tolley, N.D., Bunting, M., Schwertz, H., Jiang, H., Lindemann, S., Yost, C.C., Rubner, F.J., Albertine, K.H., Swoboda, K.J., Fratto, C.M., Tolley, E., Kraiss, L.W., McIntyre, T.M., Zimmerman, G.A. and Weyrich, A.S. (2005) Escaping the nuclear confines: signal-dependent pre-mRNA splicing in anucleate platelets. *Cell*, **122**, 379-391.

Douziech, M., Forget, D., Greenblatt, J. and Coulombe, B. (1999) Topological localization of the carboxyl-terminal domain of RNA polymerase II in the initiation complex. *J Biol Chem*, **274**, 19868-19873.

Dower, K., Kuperwasser, N., Merrikh, H. and Rosbash, M. (2004) A synthetic A tail rescues yeast nuclear accumulation of a ribozyme-terminated transcript. *RNA*, **10**, 1888-1899.

Dower, K. and Rosbash, M. (2002) T7 RNA polymerase-directed transcripts are processed in yeast and link 3' end formation to mRNA nuclear export. *RNA*, **8**, 686-697.

Dreyfuss, G., Kim, V.N. and Kataoka, N. (2002) Messenger-RNA-binding proteins and the messages they carry. *Nat Rev Mol Cell Biol*, **3**, 195-205.

Du, L. and Warren, S.L. (1997) A functional interaction between the carboxy-terminal domain of RNA polymerase II and pre-mRNA splicing. *J Cell Biol*, **136**, 5-18.

Dunn, E.F., Hammell, C.M., Hodge, C.A. and Cole, C.N. (2005) Yeast poly(A)-binding protein, Pab1, and PAN, a poly(A) nuclease complex recruited by Pab1, connect mRNA biogenesis to export. *Genes Dev*, **19**, 90-103.

Dye, M.J. and Proudfoot, N.J. (1999) Terminal exon definition occurs cotranscriptionally and promotes termination of RNA polymerase II. *Mol Cell*, **3**, 371-378.

Dye, M.J. and Proudfoot, N.J. (2001) Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell*, **105**, 669-681.

Eckner, R., Ellmeier, W. and Birnstiel, M.L. (1991) Mature mRNA 3' end formation stimulates RNA export from the nucleus. *EMBO J*, **10**, 3513-3522.

Egloff, S., O'Reilly, D., Chapman, R.D., Taylor, A., Tanzhaus, K., Pitts, L., Eick, D. and Murphy, S. (2007) Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression. *Science*, **318**, 1777-1779.

Epner, E., Reik, A., Cimbora, D., Telling, A., Bender, M.A., Fiering, S., Enver, T., Martin, D.I., Kennedy, M., Keller, G. and Groudine, M. (1998) The beta-globin LCR is not necessary for an open chromatin structure or developmentally regulated transcription of the native mouse beta-globin locus. *Mol Cell*, **2**, 447-455.

Erkmann, J.A. and Kutay, U. (2004) Nuclear export of mRNA: from the site of transcription to the cytoplasm. *Exp Cell Res*, **296**, 12-20.

Fakan, S. (1994) Perichromatin fibrils are in situ forms of nascent transcripts. *Trends Cell Biol*, **4**, 86-90.

Fakan, S., Leser, G. and Martin, T.E. (1986) Immunoelectron microscope visualization of nuclear ribonucleoprotein antigens within spread transcription complexes. *J Cell Biol*, **103**, 1153-1157.

Falck-Pedersen, E., Logan, J., Shenk, T. and Darnell, J.E., Jr. (1985) Transcription termination within the E1A gene of adenovirus induced by insertion of the mouse beta-major globin terminator element. *Cell*, **40**, 897-905.

Fan, H.Y., Merker, R.J. and Klein, H.L. (2001) High-copy-number expression of Sub2p, a member of the RNA helicase superfamily, suppresses hpr1-mediated genomic instability. *Mol Cell Biol*, **21**, 5459-5470.

Fasken, M.B. and Corbett, A.H. (2005) Process or perish: quality control in mRNA biogenesis. *Nat Struct Mol Biol*, **12**, 482-488.

Faustino, N.A. and Cooper, T.A. (2003) Pre-mRNA splicing and human disease. *Genes Dev*, **17**, 419-437.

Ferraiuolo, M.A., Lee, C.S., Ler, L.W., Hsu, J.L., Costa-Mattioli, M., Luo, M.J., Reed, R. and Sonenberg, N. (2004) A nuclear translation-like factor eIF4AIII is recruited to the mRNA during splicing and functions in nonsense-mediated decay. *Proc Natl Acad Sci U S A*, **101**, 4118-4123.

Fischer, T., Strasser, K., Racz, A., Rodriguez-Navarro, S., Oppizzi, M., Ihrig, P., Lechner, J. and Hurt, E. (2002) The mRNA export machinery requires the novel Sac3p-Thp1p complex to dock at the nucleoplasmic entrance of the nuclear pores. *EMBO J*, **21**, 5843-5852.

Fong, N. and Bentley, D.L. (2001) Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev*, **15**, 1783-1795.

Fong, N., Bird, G., Vigneron, M. and Bentley, D.L. (2003) A 10 residue motif at the C-terminus of the RNA pol II CTD is required for transcription, splicing and 3' end processing. *EMBO J*, **22**, 4274-4282.

Fong, Y.W. and Zhou, Q. (2001) Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, **414**, 929-933.

Fornerod, M., Ohno, M., Yoshida, M. and Mattaj, I.W. (1997) CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell*, **90**, 1051-1060.

Fresco, L.D. and Buratowski, S. (1996) Conditional mutants of the yeast mRNA capping enzyme show that the cap enhances, but is not required for, mRNA splicing. *RNA*, **2**, 584-596.

Galy, V., Gadal, O., Fromont-Racine, M., Romano, A., Jacquier, A. and Nehrbass, U. (2004) Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. *Cell*, **116**, 63-73.

Garrick, D., Fiering, S., Martin, D.I. and Whitelaw, E. (1998) Repeat-induced gene silencing in mammals. *Nat Genet*, **18**, 56-59.

Gatfield, D. and Izaurralde, E. (2002) REF1/Aly and the additional exon junction complex proteins are dispensable for nuclear mRNA export. *J Cell Biol*, **159**, 579-588.

Gatfield, D., Le Hir, H., Schmitt, C., Braun, I.C., Kocher, T., Wilm, M. and Izaurralde, E. (2001) The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila. *Curr Biol*, **11**, 1716-1721.

Gerber, A.P. and Keller, W. (2001) RNA editing by base deamination: more enzymes, more targets, new mysteries. *Trends Biochem Sci*, **26**, 376-384.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P. and Bickmore, W.A. (2004) Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell*, **118**, 555-566.

Gilmartin, G.M. (2005) Eukaryotic mRNA 3' processing: a common means to different ends. *Genes Dev*, **19**, 2517-2521.

Glanzer, J., Miyashiro, K.Y., Sul, J.Y., Barrett, L., Belt, B., Haydon, P. and Eberwine, J. (2005) RNA splicing capability of live neuronal dendrites. *Proc Natl Acad Sci U S A*, **102**, 16859-16864.

Gornemann, J., Kotovic, K.M., Hujer, K. and Neugebauer, K.M. (2005) Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell*, **19**, 53-63.

Gorski, S.A., Dundr, M. and Misteli, T. (2006) The road much traveled: trafficking in the cell nucleus. *Curr Opin Cell Biol*, **18**, 284-290.

Gott, J.M. and Emeson, R.B. (2000) Functions and mechanisms of RNA editing. *Annu Rev Genet*, **34**, 499-531.

Graveley, B.R. (2000) Sorting out the complexity of SR protein functions. *RNA*, **6**, 1197-1211.

Green, D.M., Johnson, C.P., Hagan, H. and Corbett, A.H. (2003) The C-terminal domain of myosin-like protein 1 (Mlp1p) is a docking site for heterogeneous nuclear ribonucleoproteins that are required for mRNA export. *Proc Natl Acad Sci U S A*, **100**, 1010-1015.

Greenleaf, A.L. (1993) Positive patches and negative noodles: linking RNA processing to transcription? *Trends Biochem Sci*, **18**, 117-119.

Gromak, N., West, S. and Proudfoot, N.J. (2006) Pause sites promote transcriptional termination of mammalian RNA polymerase II. *Mol Cell Biol*, **26**, 3986-3996.

Grosveld, F., De Boer, E., Dillon, N., Gribnau, J., Milot, E., Trimborn, T., Wijgerde, M. and Fraser, P. (1998) The dynamics of globin gene expression and gene therapy vectors. *Ann N Y Acad Sci*, **850**, 18-27.

Grosveld, F., van Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. *Cell*, **51**, 975-985.

Gruter, P., Tabernero, C., von Kobbe, C., Schmitt, C., Saavedra, C., Bachi, A., Wilm, M., Felber, B.K. and Izaurralde, E. (1998) TAP, the human homolog of Mex67p, mediates CTE-dependent RNA export from the nucleus. *Mol Cell*, **1**, 649-659.

Gu, M. and Lima, C.D. (2005) Processing the message: structural insights into capping and decapping mRNA. *Curr Opin Struct Biol*, **15**, 99-106.

Guth, S., Tange, T.O., Kellenberger, E. and Valcarcel, J. (2001) Dual function for U2AF(35) in AG-dependent pre-mRNA splicing. *Mol Cell Biol*, **21**, 7673-7681.

Guzik, B.W., Levesque, L., Prasad, S., Bor, Y.C., Black, B.E., Paschal, B.M., Rekosh, D. and Hammarskjold, M.L. (2001) NXT1 (p15) is a crucial cellular cofactor in TAP-dependent export of intron-containing RNA in mammalian cells. *Mol Cell Biol*, **21**, 2545-2554.

Hamm, J. and Mattaj, I.W. (1990) Monomethylated cap structures facilitate RNA export from the nucleus. *Cell*, **63**, 109-118.

Hammell, C.M., Gross, S., Zenklusen, D., Heath, C.V., Stutz, F., Moore, C. and Cole, C.N. (2002) Coupling of termination, 3' processing, and mRNA export. *Mol Cell Biol*, **22**, 6441-6457.

Hartner, J.C., Schmittwolf, C., Kispert, A., Muller, A.M., Higuchi, M. and Seeburg, P.H. (2004) Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. *J Biol Chem*, **279**, 4894-4902.

Hastings, M.L. and Krainer, A.R. (2001) Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol*, **13**, 302-309.

Hausman, G.M.C.a.R.E. (2004) *The Cell: A Molecular Approach*. ASM press, Washington, D.C.

Hector, R.E., Nykamp, K.R., Dheur, S., Anderson, J.T., Non, P.J., Urbinati, C.R., Wilson, S.M., Minvielle-Sebastia, L. and Swanson, M.S. (2002) Dual requirement for yeast hnRNP Nab2p in mRNA poly(A) tail length control and nuclear export. *EMBO J*, **21**, 1800-1810.

Herold, A., Klymenko, T. and Izaurralde, E. (2001) NXF1/p15 heterodimers are essential for mRNA nuclear export in Drosophila. *RNA*, **7**, 1768-1780.

Herold, A., Teixeira, L. and Izaurralde, E. (2003) Genome-wide analysis of nuclear mRNA export pathways in Drosophila. *EMBO J*, **22**, 2472-2483.

Hertel, K.J. and Graveley, B.R. (2005) RS domains contact the pre-mRNA throughout spliceosome assembly. *Trends Biochem Sci*, **30**, 115-118.

Higuchi, M., Maas, S., Single, F.N., Hartner, J., Rozov, A., Burnashev, N., Feldmeyer, D., Sprengel, R. and Seeburg, P.H. (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature*, **406**, 78-81.

Hilleren, P., McCarthy, T., Rosbash, M., Parker, R. and Jensen, T.H. (2001) Quality control of mRNA 3'-end processing is linked to the nuclear exosome. *Nature*, **413**, 538-542.

Hilleren, P. and Parker, R. (2001) Defects in the mRNA export factors Rat7p, Gle1p, Mex67p, and Rat8p cause hyperadenylation during 3'-end formation of nascent transcripts. *RNA*, **7**, 753-764.

Hirose, Y. and Manley, J.L. (1998) RNA polymerase II is an essential mRNA polyadenylation factor. *Nature*, **395**, 93-96.

Hirose, Y. and Manley, J.L. (2000) RNA polymerase II and the integration of nuclear events. *Genes Dev*, **14**, 1415-1429.

Hirose, Y., Tacke, R. and Manley, J.L. (1999) Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev*, **13**, 1234-1239.

Ho, C.K. and Shuman, S. (1999) Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Mol Cell*, **3**, 405-411.

Ho, C.K., Sriskanda, V., McCracken, S., Bentley, D., Schwer, B. and Shuman, S. (1998) The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem*, **273**, 9577-9585.

Houseley, J., LaCava, J. and Tollervey, D. (2006) RNA-quality control by the exosome. *Nat Rev Mol Cell Biol*, **7**, 529-539.

Howe, K.J. (2002) RNA polymerase II conducts a symphony of pre-mRNA processing activities. *Biochim Biophys Acta*, **1577**, 308-324.

Howe, K.J., Kane, C.M. and Ares, M., Jr. (2003) Perturbation of transcription elongation influences the fidelity of internal exon inclusion in Saccharomyces cerevisiae. *RNA*, **9**, 993-1006.

Huang, S. and Spector, D.L. (1996) Intron-dependent recruitment of pre-mRNA splicing factors to sites of transcription. *J Cell Biol*, **133**, 719-732.

Huang, Y. and Carmichael, G.G. (1996) Role of polyadenylation in nucleocytoplasmic transport of mRNA. *Mol Cell Biol*, **16**, 1534-1542.

Huang, Y., Gattoni, R., Stevenin, J. and Steitz, J.A. (2003) SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. *Mol Cell*, **11**, 837-843.

Huang, Y. and Steitz, J.A. (2001) Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol Cell*, **7**, 899-905.

Huang, Y., Yario, T.A. and Steitz, J.A. (2004) A molecular link between SR protein dephosphorylation and mRNA export. *Proc Natl Acad Sci U S A*, **101**, 9666-9670.

Iborra, F.J., Escargueil, A.E., Kwek, K.Y., Akoulitchev, A. and Cook, P.R. (2004) Molecular cross-talk between the transcription, translation, and nonsense-mediated decay machineries. *J Cell Sci*, **117**, 899-906.

Iborra, F.J., Jackson, D.A. and Cook, P.R. (2001) Coupled Transcription and Translation Within Nuclei of Mammalian Cells. *Science*, **21**, 21.

Iborra, F.J., Pombo, A., McManus, J., Jackson, D.A. and Cook, P.R. (1996) The topology of transcription by immobilized polymerases. *Exp Cell Res*, **229**, 167-173.

Izaurralde, E. (2002) A novel family of nuclear transport receptors mediates the export of messenger RNA to the cytoplasm. *Eur J Cell Biol*, **81**, 577-584.

Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E. and Mattaj, I.W. (1994) A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell*, **78**, 657-668.

Jackson, D.A., Bartlett, J. and Cook, P.R. (1996) Sequences attaching loops of nuclear and mitochondrial DNA to underlying structures in human cells: the role of transcription units. *Nucleic Acids Res*, **24**, 1212-1219.

Jackson, R.J. (2005) Alternative mechanisms of initiating translation of mammalian mRNAs. *Biochem Soc Trans*, **33**, 1231-1241.

Jensen, T.H., Boulay, J., Rosbash, M. and Libri, D. (2001a) The DECD box putative ATPase Sub2p is an early mRNA export factor. *Curr Biol*, **11**, 1711-1715.

Jensen, T.H., Dower, K., Libri, D. and Rosbash, M. (2003) Early formation of mRNP: license for export or quality control? *Mol Cell*, **11**, 1129-1138.

Jensen, T.H., Patricio, K., McCarthy, T. and Rosbash, M. (2001b) A block to mRNA nuclear export in S. cerevisiae leads to hyperadenylation of transcripts that accumulate at the site of transcription. *Mol Cell*, **7**, 887-898.

Jimeno, S., Rondon, A.G., Luna, R. and Aguilera, A. (2002) The yeast THO complex and mRNA export factors link RNA metabolism with transcription and genome instability. *EMBO J*, **21**, 3526-3535.

Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell*, **12**, 5-14.

Katahira, J., Strasser, K., Podtelejnikov, A., Mann, M., Jung, J.U. and Hurt, E. (1999) The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. *EMBO J*, **18**, 2593-2609.

Kataoka, N., Diem, M.D., Kim, V.N., Yong, J. and Dreyfuss, G. (2001) Magoh, a human homolog of Drosophila mago nashi protein, is a component of the splicing-dependent exon-exon junction complex. *EMBO J*, **20**, 6424-6433.

Kataoka, N., Yong, J., Kim, V.N., Velazquez, F., Perkinson, R.A., Wang, F. and Dreyfuss, G. (2000) Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol Cell*, **6**, 673-682.

Keller, R.W., Kuhn, U., Aragon, M., Bornikova, L., Wahle, E. and Bear, D.G. (2000) The nuclear poly(A) binding protein, PABP2, forms an oligomeric particle covering the length of the poly(A) tail. *J Mol Biol*, **297**, 569-583.

Kent, O.A., Reayi, A., Foong, L., Chilibeck, K.A. and MacMillan, A.M. (2003) Structuring of the 3' splice site by U2AF65. *J Biol Chem*, **278**, 50572-50577.

Khajavi, M., Inoue, K. and Lupski, J.R. (2006) Nonsense-mediated mRNA decay modulates clinical outcome of genetic disease. *Eur J Hum Genet*, **14**, 1074-1081.

Kiesler, E., Miralles, F. and Visa, N. (2002) HEL/UAP56 binds cotranscriptionally to the Balbiani ring pre-mRNA in an intron-independent manner and accompanies the BR mRNP to the nuclear pore. *Curr Biol*, **12**, 859-862.

Kim, E., Du, L., Bregman, D.B. and Warren, S.L. (1997) Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA. *J Cell Biol*, **136**, 19-28.

Kim, H.J., Jeong, S.H., Heo, J.H., Jeong, S.J., Kim, S.T., Youn, H.D., Han, J.W., Lee, H.W. and Cho, E.J. (2004a) mRNA capping enzyme activity is coupled to an early transcription elongation. *Mol Cell Biol*, **24**, 6184-6193.

Kim, M., Krogan, N.J., Vasiljeva, L., Rando, O.J., Nedea, E., Greenblatt, J.F. and Buratowski, S. (2004b) The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, **432**, 517-522.

Kim, W.Y. and Dahmus, M.E. (1986) Immunochemical analysis of mammalian RNA polymerase II subspecies. Stability and relative in vivo concentration. *J Biol Chem*, **261**, 14219-14225.

Kiseleva, E., Wurtz, T., Visa, N. and Daneholt, B. (1994) Assembly and disassembly of spliceosomes along a specific pre-messenger RNP fiber. *EMBO J*, **13**, 6052-6061.

Kobor, M.S., Archambault, J., Lester, W., Holstege, F.C., Gileadi, O., Jansma, D.B., Jennings, E.G., Kouyoumdjian, F., Davidson, A.R., Young, R.A. and Greenblatt, J. (1999) An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae. *Mol Cell*, **4**, 55-62.

Kobor, M.S. and Greenblatt, J. (2002) Regulation of transcription elongation by phosphorylation. *Biochim Biophys Acta*, **1577**, 261-275.

Komarnitsky, P., Cho, E.J. and Buratowski, S. (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev*, **14**, 2452-2460.

Konig, H., Matter, N., Bader, R., Thiele, W. and Muller, F. (2007) Splicing segregation: the minor spliceosome acts outside the nucleus and controls cell proliferation. *Cell*, **131**, 718-729.

Kornblihtt, A.R. (2005) Promoter usage and alternative splicing. *Curr Opin Cell Biol*, **17**, 262-268.

Kornblihtt, A.R., de la Mata, M., Fededa, J.P., Munoz, M.J. and Nogues, G. (2004) Multiple links between transcription and splicing. *RNA*, **10**, 1489-1498.

Kuai, L., Das, B. and Sherman, F. (2005) A nuclear degradation pathway controls the abundance of normal mRNAs in Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A*, **102**, 13962-13967.

Kudo, N., Matsumori, N., Taoka, H., Fujiwara, D., Schreiner, E.P., Wolff, B., Yoshida, M. and Horinouchi, S. (1999) Leptomycin B inactivates CRM1/exportin 1 by covalent modification at a cysteine residue in the central conserved region. *Proc Natl Acad Sci U S A*, **96**, 9112-9117.

Kuersten, S. and Goodwin, E.B. (2005) Linking nuclear mRNP assembly and cytoplasmic destiny. *Biol Cell*, **97**, 469-478.

Kuhn, U., Nemeth, A., Meyer, S. and Wahle, E. (2003) The RNA binding domains of the nuclear poly(A)-binding protein. *J Biol Chem*, **278**, 16916-16925.

Kuhn, U. and Wahle, E. (2004) Structure and function of poly(A) binding proteins. *Biochim Biophys Acta*, **1678**, 67-84.

LaCava, J., Houseley, J., Saveanu, C., Petfalski, E., Thompson, E., Jacquier, A. and Tollervey, D. (2005) RNA degradation by the exosome is promoted by a nuclear polyadenylation complex. *Cell*, **121**, 713-724.

Lamond, A.I., Konarska, M.M. and Sharp, P.A. (1987) A mutational analysis of spliceosome assembly: evidence for splice site collaboration during spliceosome formation. *Genes Dev*, **1**, 532-543.

Lamond, A.I. and Spector, D.L. (2003) Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol*, **4**, 605-612.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., et al. (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860-921.

Laurencikiene, J., Kallman, A.M., Fong, N., Bentley, D.L. and Ohman, M. (2006) RNA editing and alternative splicing: the importance of co-transcriptional coordination. *EMBO Rep*, **7**, 303-307.

Le Hir, H., Gatfield, D., Izaurralde, E. and Moore, M.J. (2001) The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. *EMBO J*, **20**, 4987-4997.

Le Hir, H., Izaurralde, E., Maquat, L.E. and Moore, M.J. (2000a) The spliceosome deposits multiple proteins 20-24 nucleotides upstream of mRNA exon-exon junctions. *EMBO J*, **19**, 6860-6869.

Le Hir, H., Moore, M.J. and Maquat, L.E. (2000b) Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev*, **14**, 1098-1108.

Legrain, P. and Rosbash, M. (1989) Some cis- and trans-acting mutants for splicing target pre-mRNA to the cytoplasm. *Cell*, **57**, 573-583.

Lei, E.P., Krebber, H. and Silver, P.A. (2001) Messenger RNAs are recruited for nuclear export during transcription. *Genes Dev*, **15**, 1771-1782.

Lei, E.P. and Silver, P.A. (2002a) Intron status and 3'-end formation control cotranscriptional export of mRNA. *Genes Dev*, **16**, 2761-2766.

Lei, E.P. and Silver, P.A. (2002b) Protein and RNA export from the nucleus. *Dev Cell*, **2**, 261-272.

Lei, E.P., Stern, C.A., Fahrenkrog, B., Krebber, H., Moy, T.I., Aebi, U. and Silver, P.A. (2003) Sac3 is an mRNA export factor that localizes to cytoplasmic fibrils of nuclear pore complex. *Mol Biol Cell*, **14**, 836-847.

Lejeune, F., Ishigaki, Y., Li, X. and Maquat, L.E. (2002) The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: dynamics of mRNP remodeling. *EMBO J*, **21**, 3536-3545.

Levings, P.P. and Bungert, J. (2002) The human beta-globin locus control region. *Eur J Biochem*, **269**, 1589-1599.

Lewis, J.D., Izaurralde, E., Jarmolowski, A., McGuigan, C. and Mattaj, I.W. (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev*, **10**, 1683-1698.

Li, C., Lin, R.I., Lai, M.C., Ouyang, P. and Tarn, W.Y. (2003) Nuclear Pnn/DRS protein binds to spliced mRNPs and participates in mRNA processing and export via interaction with RNPS1. *Mol Cell Biol*, **23**, 7363-7376.

Li, Q., Peterson, K.R., Fang, X. and Stamatoyannopoulos, G. (2002) Locus control regions. *Blood*, **100**, 3077-3086.

Libri, D., Dower, K., Boulay, J., Thomsen, R., Rosbash, M. and Jensen, T.H. (2002) Interactions between mRNA export commitment, 3'-end quality control, and nuclear degradation. *Mol Cell Biol*, **22**, 8254-8266.

Libri, D., Graziani, N., Saguez, C. and Boulay, J. (2001) Multiple roles for the yeast SUB2/yUAP56 gene in splicing. *Genes Dev*, **15**, 36-41.

Logan, J., Falck-Pedersen, E., Darnell, J.E., Jr. and Shenk, T. (1987) A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proc Natl Acad Sci U S A*, **84**, 8306-8310.

Long, R.M., Elliott, D.J., Stutz, F., Rosbash, M. and Singer, R.H. (1995) Spatial consequences of defective processing of specific yeast mRNAs revealed by fluorescent in situ hybridization. *RNA*, **1**, 1071-1078.

Longman, D., Johnstone, I.L. and Caceres, J.F. (2003) The Ref/Aly proteins are dispensable for mRNA export and development in Caenorhabditis elegans. *RNA*, **9**, 881-891.

Lu, H., Flores, O., Weinmann, R. and Reinberg, D. (1991) The nonphosphorylated form of RNA polymerase II preferentially associates with the preinitiation complex. *Proc Natl Acad Sci U S A*, **88**, 10004-10008.

Luo, M.J. and Reed, R. (1999) Splicing is required for rapid and efficient mRNA export in metazoans. *Proc Natl Acad Sci U S A*, **96**, 14937-14942.

Luo, M.L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M. and Reed, R. (2001) Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature*, **413**, 644-647.

Luo, W., Johnson, A.W. and Bentley, D.L. (2006) The role of Rat1 in coupling mRNA 3'-end processing to transcription termination: implications for a unified allosteric-torpedo model. *Genes Dev*, **20**, 954-965.

Lykke-Andersen, J., Shu, M.D. and Steitz, J.A. (2001) Communication of the position of exon-exon junctions to the mRNA surveillance machinery by the protein RNPS1. *Science*, **293**, 1836-1839.

MacMorris, M., Brocker, C. and Blumenthal, T. (2003) UAP56 levels affect viability and mRNA export in Caenorhabditis elegans. *RNA*, **9**, 847-857.

Mandal, S.S., Chu, C., Wada, T., Handa, H., Shatkin, A.J. and Reinberg, D. (2004) Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc Natl Acad Sci U S A*, **101**, 7572-7577.

Mangus, D.A., Evans, M.C. and Jacobson, A. (2003) Poly(A)-binding proteins: multifunctional scaffolds for the post-transcriptional control of gene expression. *Genome Biol*, **4**, 223.

Maniatis, T. and Reed, R. (2002) An extensive network of coupling among gene expression machines. *Nature*, **416**, 499-506.

Masuda, S., Das, R., Cheng, H., Hurt, E., Dorman, N. and Reed, R. (2005) Recruitment of the human TREX complex to mRNA during splicing. *Genes Dev*, **19**, 1512-1517.

Matlin, A.J., Clark, F. and Smith, C.W. (2005) Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol*, **6**, 386-398.

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S. and Bentley, D.L. (1997a) 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes Dev*, **11**, 3306-3318.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M. and Bentley, D.L. (1997b) The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, **385**, 357-361.

Meaburn, K.J. and Misteli, T. (2007) Cell biology: chromosome territories. *Nature*, **445**, 379-781.

Meinhart, A. and Cramer, P. (2004) Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors. *Nature*, **430**, 223-226.

Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S. and Cramer, P. (2005) A structural perspective of CTD function. *Genes Dev*, **19**, 1401-1415.

Meininghaus, M., Chapman, R.D., Horndasch, M. and Eick, D. (2000) Conditional expression of RNA polymerase II in mammalian cells. Deletion of the carboxyl-terminal domain of the large subunit affects early steps in transcription. *J Biol Chem*, **275**, 24375-24382.

Millevoi, S., Geraghty, F., Idowu, B., Tam, J.L., Antoniou, M. and Vagner, S. (2002) A novel function for the U2AF 65 splicing factor in promoting pre-mRNA 3'-end processing. *EMBO Rep*, **3**, 869-874.

Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M. and Vagner, S. (2006) An interaction between U2AF 65 and CF I(m) links the splicing and 3' end processing machineries. *EMBO J*, **25**, 4854-4864.

Millhouse, S. and Manley, J.L. (2005) The C-terminal domain of RNA polymerase II functions as a phosphorylation-dependent splicing activator in a heterologous protein. *Mol Cell Biol*, **25**, 533-544.

Misteli, T. (2000) Cell biology of transcription and pre-mRNA splicing: nuclear architecture meets nuclear function. *J Cell Sci*, **113 (Pt 11)**, 1841-1849.

Misteli, T. (2005) Concepts in nuclear architecture. *Bioessays*, **27**, 477-487.
Misteli, T. (2008) Physiological importance of RNA and protein mobility in the cell nucleus. *Histochem Cell Biol*, **129**, 5-11.
Misteli, T., Caceres, J.F. and Spector, D.L. (1997) The dynamics of a pre-mRNA splicing factor in living cells. *Nature*, **387**, 523-527.
Misteli, T. and Spector, D.L. (1998) The cellular organization of gene expression. *Curr Opin Cell Biol*, **10**, 323-331.
Misteli, T. and Spector, D.L. (1999) RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo. *Mol Cell*, **3**, 697-705.
Mitchell, P. and Tollervey, D. (2001) mRNA turnover. *Curr Opin Cell Biol*, **13**, 320-325.
Modrek, B. and Lee, C. (2002) A genomic view of alternative splicing. *Nat Genet*, **30**, 13-19.
Molenaar, C., Abdulle, A., Gena, A., Tanke, H.J. and Dirks, R.W. (2004) Poly(A)+ RNAs roam the cell nucleus and pass through speckle domains in transcriptionally active and inactive cells. *J Cell Biol*, **165**, 191-202.
Moore, M.J. (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science*, **309**, 1514-1518.
Mortillaro, M.J., Blencowe, B.J., Wei, X., Nakayasu, H., Du, L., Warren, S.L., Sharp, P.A. and Berezney, R. (1996) A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix. *Proc Natl Acad Sci U S A*, **93**, 8253-8257.
Moteki, S. and Price, D. (2002) Functional coupling of capping and transcription of mRNA. *Mol Cell*, **10**, 599-609.
Myers, L.C., Lacomis, L., Erdjument-Bromage, H. and Tempst, P. (2002) The yeast capping enzyme represses RNA polymerase II transcription. *Mol Cell*, **10**, 883-894.
Needham, M., Gooding, C., Hudson, K., Antoniou, M., Grosveld, F. and Hollis, M. (1992) LCR/MEL: a versatile system for high-level expression of heterologous proteins in erythroid cells. *Nucleic Acids Res*, **20**, 997-1003.
Niwa, M. and Berget, S.M. (1991) Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev*, **5**, 2086-2095.
Niwa, M., Rose, S.D. and Berget, S.M. (1990) In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev*, **4**, 1552-1559.
Nogues, G., Kadener, S., Cramer, P., de la Mata, M., Fededa, J.P., Blaustein, M., Srebrow, A. and Kornblihtt, A.R. (2003) Control of alternative pre-mRNA splicing by RNA Pol II elongation: faster is not always better. *IUBMB Life*, **55**, 235-241.
O'Brien, T., Hardin, S., Greenleaf, A. and Lis, J.T. (1994) Phosphorylation of RNA polymerase II C-terminal domain and transcriptional elongation. *Nature*, **370**, 75-77.
O'Keefe, R.T., Mayeda, A., Sadowski, C.L., Krainer, A.R. and Spector, D.L. (1994) Disruption of pre-mRNA splicing in vivo results in reorganization of splicing factors. *J Cell Biol*, **124**, 249-260.
O'Mullane, L. and Eperon, I.C. (1998) The pre-mRNA 5' cap determines whether U6 small nuclear RNA succeeds U1 small nuclear ribonucleoprotein particle at 5' splice sites. *Mol Cell Biol*, **18**, 7510-7520.
O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J. and Proudfoot, N.J. (2004) Gene loops juxtapose promoters and terminators in yeast. *Nat Genet*, **36**, 1014-1018.
Ohno, M., Segref, A., Kuersten, S. and Mattaj, I.W. (2002) Identity elements used in export of mRNAs. *Mol Cell*, **9**, 659-671.
Olivieri, N.F. (1999) The beta-thalassemias. *N Engl J Med*, **341**, 99-109.
Orphanides, G. and Reinberg, D. (2002) A unified theory of gene expression. *Cell*, **108**, 439-451.
Osheim, Y.N., Miller, O.L., Jr. and Beyer, A.L. (1988) Visualization of Drosophila melanogaster chorion genes undergoing amplification. *Mol Cell Biol*, **8**, 2811-2821.
Osheim, Y.N., Proudfoot, N.J. and Beyer, A.L. (1999) EM visualization of transcription by RNA polymerase II: downstream termination requires a poly(A) signal but not transcript cleavage. *Mol Cell*, **3**, 379-387.
Osheim, Y.N., Sikes, M.L. and Beyer, A.L. (2002) EM visualization of Pol II genes in Drosophila: most genes terminate without prior 3' end cleavage of nascent transcripts. *Chromosoma*, **111**, 1-12.
Pacheco, T.R., Coelho, M.B., Desterro, J.M., Mollet, I. and Carmo-Fonseca, M. (2006) In vivo requirement of the small subunit of U2AF for recognition of a weak 3' splice site. *Mol Cell Biol*, **26**, 8183-8190.
Pacheco, T.R., Gomes, A.Q., Barbosa-Morais, N.L., Benes, V., Ansorge, W., Wollerton, M., Smith, C.W., Valcarcel, J. and Carmo-Fonseca, M. (2004) Diversity of vertebrate splicing factor U2AF35: identification of alternatively spliced U2AF1 mRNAS. *J Biol Chem*, **279**, 27039-27049.
Palancade, B. and Bensaude, O. (2003) Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur J Biochem*, **270**, 3859-3870.
Pei, Y. and Shuman, S. (2002) Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J Biol Chem*, **277**, 19639-19648.

Phatnani, H.P. and Greenleaf, A.L. (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev*, **20**, 2922-2936.

Pillutla, R.C., Yue, Z., Maldonado, E. and Shatkin, A.J. (1998) Recombinant human mRNA cap methyltransferase binds capping enzyme/RNA polymerase IIo complexes. *J Biol Chem*, **273**, 21443-21446.

Politz, J.C., Browne, E.S., Wolf, D.E. and Pederson, T. (1998) Intranuclear diffusion and hybridization state of oligonucleotides measured by fluorescence correlation spectroscopy in living cells. *Proc Natl Acad Sci U S A*, **95**, 6043-6048.

Politz, J.C., Tuft, R.A., Pederson, T. and Singer, R.H. (1999) Movement of nuclear poly(A) RNA throughout the interchromatin space in living cells. *Curr Biol*, **9**, 285-291.

Pombo, A., Jones, E., Iborra, F.J., Kimura, H., Sugaya, K., Cook, P.R. and Jackson, D.A. (2000) Specialized transcription factories within mammalian nuclei. *Crit Rev Eukaryot Gene Expr*, **10**, 21-29.

Powell, L.M., Wallis, S.C., Pease, R.J., Edwards, Y.H., Knott, T.J. and Scott, J. (1987) A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell*, **50**, 831-840.

Proudfoot, N. (2004) New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol*, **16**, 272-278.

Proudfoot, N. and O'Sullivan, J. (2002) Polyadenylation: a tail of two complexes. *Curr Biol*, **12**, R855-857.

Proudfoot, N.J. (1989) How RNA polymerase II terminates transcription in higher eukaryotes. *Trends Biochem Sci*, **14**, 105-110.

Proudfoot, N.J., Furger, A. and Dye, M.J. (2002) Integrating mRNA processing with transcription. *Cell*, **108**, 501-512.

Rain, J.C. and Legrain, P. (1997) In vivo commitment to splicing in yeast involves the nucleotide upstream from the branch site conserved sequence and the Mud2 protein. *EMBO J*, **16**, 1759-1771.

Rasmussen, E.B. and Lis, J.T. (1993) In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proc Natl Acad Sci U S A*, **90**, 7923-7927.

Reed, R. (2003) Coupling transcription, splicing and mRNA export. *Curr Opin Cell Biol*, **15**, 326-331.

Reed, R. and Cheng, H. (2005) TREX, SR proteins and export of mRNA. *Curr Opin Cell Biol*, **17**, 269-273.

Reed, R. and Hurt, E. (2002) A conserved mRNA export machinery coupled to pre-mRNA splicing. *Cell*, **108**, 523-531.

Rehwinkel, J., Herold, A., Gari, K., Kocher, T., Rode, M., Ciccarelli, F.L., Wilm, M. and Izaurralde, E. (2004) Genome-wide analysis of mRNAs regulated by the THO complex in Drosophila melanogaster. *Nat Struct Mol Biol*, **11**, 558-566.

Reich, E. and Goldberg, I.H. (1964) Actinomycin and nucleic acid function. *Prog Nucleic Acid Res Mol Biol*, **3**, 183-234.

Reichert, V.L., Le Hir, H., Jurica, M.S. and Moore, M.J. (2002) 5' exon interactions within the human spliceosome establish a framework for exon junction complex structure and assembly. *Genes Dev*, **16**, 2778-2791.

Reik, A., Telling, A., Zitnik, G., Cimbora, D., Epner, E. and Groudine, M. (1998) The locus control region is necessary for gene expression in the human beta-globin locus but not the maintenance of an open chromatin structure in erythroid cells. *Mol Cell Biol*, **18**, 5992-6000.

Robberson, B.L., Cote, G.J. and Berget, S.M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol*, **10**, 84-94.

Rodrigues, J.P., Rode, M., Gatfield, D., Blencowe, B.J., Carmo-Fonseca, M. and Izaurralde, E. (2001) REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus. *Proc Natl Acad Sci U S A*, **98**, 1030-1035.

Rodriguez, C.R., Cho, E.J., Keogh, M.C., Moore, C.L., Greenleaf, A.L. and Buratowski, S. (2000) Kin28, the TFIIH-associated carboxy-terminal domain kinase, facilitates the recruitment of mRNA processing machinery to RNA polymerase II. *Mol Cell Biol*, **20**, 104-112.

Rondón, A.G., Jimeno, S., Garcia-Rubio, M. and Aguilera, A. (2003) Molecular evidence that the eukaryotic THO/TREX complex is required for efficient transcription elongation. *J Biol Chem*, **278**, 39037-39043.

Rosonina, E. and Blencowe, B.J. (2002) Gene expression: the close coupling of transcription and splicing. *Curr Biol*, **12**, R319-321.

Rosonina, E., Kaneko, S. and Manley, J.L. (2006) Terminating the transcript: breaking up is hard to do. *Genes Dev*, **20**, 1050-1056.

Rougemaille, M., Gudipati, R.K., Olesen, J.R., Thomsen, R., Seraphin, B., Libri, D. and Jensen, T.H. (2007) Dissecting mechanisms of nuclear mRNA surveillance in THO/sub2 complex mutants. *Embo J*, **26**, 2317-2326.

Ryan, K., Calvo, O. and Manley, J.L. (2004) Evidence that polyadenylation factor CPSF-73 is the mRNA 3' processing endonuclease. *RNA*, **10**, 565-573.

Ryan, K., Murthy, K.G., Kaneko, S. and Manley, J.L. (2002) Requirements of the RNA polymerase II C-terminal domain for reconstituting pre-mRNA 3' cleavage. *Mol Cell Biol*, **22**, 1684-1692.

Ryan, K.J. and Wente, S.R. (2000) The nuclear pore complex: a protein machine bridging the nucleus and cytoplasm. *Curr Opin Cell Biol*, **12**, 361-371.

Ryman, K., Fong, N., Bratt, E., Bentley, D.L. and Ohman, M. (2007) The C-terminal domain of RNA Pol II helps ensure that editing precedes splicing of the GluR-B transcript. *RNA*, **13**, 1071-1078.

Sadowski, M., Dichtl, B., Hubner, W. and Keller, W. (2003) Independent functions of yeast Pcf11p in pre-mRNA 3' end processing and in transcription termination. *EMBO J*, **22**, 2167-2177.

Saguez, C., Olesen, J.R. and Jensen, T.H. (2005) Formation of export-competent mRNP: escaping nuclear destruction. *Curr Opin Cell Biol*, **17**, 287-293.

Sakashita, E., Tatsumi, S., Werner, D., Endo, H. and Mayeda, A. (2004) Human RNPS1 and its associated factors: a versatile alternative pre-mRNA splicing regulator in vivo. *Mol Cell Biol*, **24**, 1174-1187.

Sanford, J.R. and Caceres, J.F. (2004) Pre-mRNA splicing: life at the centre of the central dogma. *J Cell Sci*, **117**, 6261-6263.

Santos-Rosa, H., Moreno, H., Simos, G., Segref, A., Fahrenkrog, B., Pante, N. and Hurt, E. (1998) Nuclear mRNA export requires complex formation between Mex67p and Mtr2p at the nuclear pores. *Mol Cell Biol*, **18**, 6826-6838.

Schaub, M. and Keller, W. (2002) RNA editing by adenosine deaminases generates RNA and protein diversity. *Biochimie*, **84**, 791-803.

Schilders, G., van Dijk, E., Raijmakers, R. and Pruijn, G.J. (2006) Cell and molecular biology of the exosome: how to make or break an RNA. *Int Rev Cytol*, **251**, 159-208.

Schneiter, R., Guerra, C.E., Lampl, M., Gogg, G., Kohlwein, S.D. and Klein, H.L. (1999) The Saccharomyces cerevisiae hyperrecombination mutant hpr1Delta is synthetically lethal with two conditional alleles of the acetyl coenzyme A carboxylase gene and causes a defect in nuclear export of polyadenylated RNA. *Mol Cell Biol*, **19**, 3415-3422.

Schroeder, S.C., Schwer, B., Shuman, S. and Bentley, D. (2000) Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev*, **14**, 2435-2440.

Schroeder, S.C., Zorio, D.A., Schwer, B., Shuman, S. and Bentley, D. (2004) A function of yeast mRNA cap methyltransferase, Abd1, in transcription by RNA polymerase II. *Mol Cell*, **13**, 377-387.

Schwer, B. and Shuman, S. (1996) Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing in vivo. *RNA*, **2**, 574-583.

Schwerk, C., Prasad, J., Degenhardt, K., Erdjument-Bromage, H., White, E., Tempst, P., Kidd, V.J., Manley, J.L., Lahti, J.M. and Reinberg, D. (2003) ASAP, a novel protein complex involved in RNA processing and apoptosis. *Mol Cell Biol*, **23**, 2981-2990.

Segref, A., Sharma, K., Doye, V., Hellwig, A., Huber, J., Luhrmann, R. and Hurt, E. (1997) Mex67p, a novel factor for nuclear mRNA export, binds to both poly(A)+ RNA and nuclear pores. *EMBO J*, **16**, 3256-3271.

Seksek, O., Biwersi, J. and Verkman, A.S. (1997) Translational diffusion of macromolecule-sized solutes in cytoplasm and nucleus. *J Cell Biol*, **138**, 131-142.

Selenko, P., Gregorovic, G., Sprangers, R., Stier, G., Rhani, Z., Kramer, A. and Sattler, M. (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell*, **11**, 965-976.

Sharp, P.A. (2005) The discovery of split genes and RNA splicing. *Trends Biochem Sci*, **30**, 279-281.

Shav-Tal, Y., Darzacq, X., Shenoy, S.M., Fusco, D., Janicki, S.M., Spector, D.L. and Singer, R.H. (2004) Dynamics of single mRNPs in nuclei of living cells. *Science*, **304**, 1797-1800.

Shav-Tal, Y., Darzacq, X. and Singer, R.H. (2006) Gene expression within a dynamic nuclear landscape. *EMBO J*, **25**, 3469-3479.

Shen, H. and Green, M.R. (2004) A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol Cell*, **16**, 363-373.

Shibuya, T., Tange, T.O., Sonenberg, N. and Moore, M.J. (2004) eIF4AIII binds spliced mRNA in the exon junction complex and is essential for nonsense-mediated decay. *Nat Struct Mol Biol*, **11**, 346-351.

Shilatifard, A., Conaway, R.C. and Conaway, J.W. (2003) The RNA polymerase II elongation complex. *Annu Rev Biochem*, **72**, 693-715.

Shuman, S. (2001) Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog Nucleic Acid Res Mol Biol*, **66**, 1-40.

Sisodia, S.S., Sollner-Webb, B. and Cleveland, D.W. (1987) Specificity of RNA maturation pathways: RNAs transcribed by RNA polymerase III are not substrates for splicing or polyadenylation. *Mol Cell Biol*, **7**, 3602-3612.

Smale, S.T. and Tjian, R. (1985) Transcription of herpes simplex virus tk sequences under the control of wild-type and mutant human RNA polymerase I promoters. *Mol Cell Biol*, **5**, 352-362.
Snaar, S.P., Verdijk, P., Tanke, H.J. and Dirks, R.W. (2002) Kinetics of HCMV immediate early mRNA expression in stably transfected fibroblasts. *J Cell Sci*, **115**, 321-328.
Sommer, P. and Nehrbass, U. (2005) Quality control of messenger ribonucleoprotein particles in the nucleus and at the pore. *Curr Opin Cell Biol*, **17**, 294-301.
Spector, D.L. (1993) Macromolecular domains within the cell nucleus. *Annu Rev Cell Biol*, **9**, 265-315.
Spector, D.L., Schrier, W.H. and Busch, H. (1983) Immunoelectron microscopic localization of snRNPs. *Biol Cell*, **49**, 1-10.
Stewart, M. (2007) Ratcheting mRNA out of the nucleus. *Mol Cell*, **25**, 327-330.
Stiller, J.W. and Hall, B.D. (2002) Evolution of the RNA polymerase II C-terminal domain. *Proc Natl Acad Sci U S A*, **99**, 6091-6096.
Strasser, K. and Hurt, E. (2000) Yra1p, a conserved nuclear RNA-binding protein, interacts directly with Mex67p and is required for mRNA export. *EMBO J*, **19**, 410-420.
Strasser, K. and Hurt, E. (2001) Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature*, **413**, 648-652.
Strasser, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondon, A.G., Aguilera, A., Struhl, K., Reed, R. and Hurt, E. (2002) TREX is a conserved complex coupling transcription with messenger RNA export. *Nature*, **417**, 304-308.
Stutz, F., Bachi, A., Doerks, T., Braun, I.C., Seraphin, B., Wilm, M., Bork, P. and Izaurralde, E. (2000) REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA*, **6**, 638-650.
Stutz, F. and Izaurralde, E. (2003) The interplay of nuclear mRNP assembly, mRNA surveillance and export. *Trends Cell Biol*, **13**, 319-327.
Sugaya, K., Vigneron, M. and Cook, P.R. (2000) Mammalian cell lines expressing functional RNA polymerase II tagged with the green fluorescent protein. *J Cell Sci*, **113 (Pt 15)**, 2679-2683.
Szentirmay, M.N. and Sawadogo, M. (2000) Spatial organization of RNA polymerase II transcription in the nucleus. *Nucleic Acids Res*, **28**, 2019-2025.
Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosveld, F. and Greaves, D.R. (1989) A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. *Nature*, **338**, 352-355.
Tang, Y., Liu, D.P. and Liang, C.C. (2002) Further understanding of the beta-globin locus regulation at the molecular level: looping or linking models? *Genes Cells*, **7**, 889-900.
Tange, T.O., Nott, A. and Moore, M.J. (2004) The ever-increasing complexities of the exon junction complex. *Curr Opin Cell Biol*, **16**, 279-284.
Tange, T.O., Shibuya, T., Jurica, M.S. and Moore, M.J. (2005) Biochemical analysis of the EJC reveals two new factors and a stable tetrameric protein core. *RNA*, **11**, 1869-1883.
Tavanez, J.P., Calado, P., Braga, J., Lafarga, M. and Carmo-Fonseca, M. (2005) In vivo aggregation properties of the nuclear poly(A)-binding protein PABPN1. *RNA*, **11**, 752-762.
Teixeira, A., Tahiri-Alaoui, A., West, S., Thomas, B., Ramadass, A., Martianov, I., Dye, M., James, W., Proudfoot, N.J. and Akoulitchev, A. (2004) Autocatalytic RNA cleavage in the human beta-globin pre-mRNA promotes transcription termination. *Nature*, **432**, 526-530.
Thomsen, R., Libri, D., Boulay, J., Rosbash, M. and Jensen, T.H. (2003) Localization of nuclear retained mRNAs in Saccharomyces cerevisiae. *RNA*, **9**, 1049-1057.
Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F. and de Laat, W. (2002) Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*, **10**, 1453-1465.
Tollervey, D. (2004) Molecular biology: termination by torpedo. *Nature*, **432**, 456-457.
Tsiftsoglou, A.S., Pappas, I.S. and Vizirianakis, I.S. (2003) Mechanisms involved in the induced differentiation of leukemia cells. *Pharmacol Ther*, **100**, 257-290.
Tuan, D., Kong, S. and Hu, K. (1992) Transcription of the hypersensitive site HS2 enhancer in erythroid cells. *Proc Natl Acad Sci U S A*, **89**, 11219-11223.
Vagner, S., Vagner, C. and Mattaj, I.W. (2000) The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev*, **14**, 403-413.
Vanacova, S. and Stefl, R. (2007) The exosome and RNA quality control in the nucleus. *EMBO Rep*, **8**, 651-657.
Vargas, D.Y., Raj, A., Marras, S.A., Kramer, F.R. and Tyagi, S. (2005) Mechanism of mRNA transport in the nucleus. *Proc Natl Acad Sci U S A*, **102**, 17008-17013.
Vasu, S.K. and Forbes, D.J. (2001) Nuclear pores and nuclear assembly. *Curr Opin Cell Biol*, **13**, 363-375.
Vasudevan, S. and Peltz, S.W. (2003) Nuclear mRNA surveillance. *Curr Opin Cell Biol*, **15**, 332-337.

Vinciguerra, P., Iglesias, N., Camblong, J., Zenklusen, D. and Stutz, F. (2005) Perinuclear Mlp proteins downregulate gene expression in response to a defect in mRNA export. *EMBO J*, **24**, 813-823.

Wen, Y. and Shatkin, A.J. (1999) Transcription elongation factor hSPT5 stimulates mRNA capping. *Genes Dev*, **13**, 1774-1779.

West, M.L. and Corden, J.L. (1995) Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. *Genetics*, **140**, 1223-1233.

West, S., Gromak, N., Norbury, C.J. and Proudfoot, N.J. (2006) Adenylation and exosome-mediated degradation of cotranscriptionally cleaved pre-messenger RNA in human cells. *Mol Cell*, **21**, 437-443.

West, S., Gromak, N. and Proudfoot, N.J. (2004) Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, **432**, 522-525.

West, S. and Proudfoot, N.J. (2007) Human Pcf11 enhances degradation of RNA polymerase II-associated nascent RNA and transcriptional termination. *Nucleic Acids Res*.

Whitelaw, E. and Proudfoot, N. (1986) Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *EMBO J*, **5**, 2915-2922.

Wijgerde, M., Grosveld, F. and Fraser, P. (1995) Transcription complex stability and chromatin dynamics in vivo. *Nature*, **377**, 209-213.

Will, C.L. and Luhrmann, R. (2005) Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol Chem*, **386**, 713-724.

Woychik, N.A. and Hampsey, M. (2002) The RNA polymerase II machinery: structure illuminates function. *Cell*, **108**, 453-463.

Xu, Y.X., Hirose, Y., Zhou, X.Z., Lu, K.P. and Manley, J.L. (2003) Pin1 modulates the structure and function of human RNA polymerase II. *Genes Dev*, **17**, 2765-2776.

Yonaha, M. and Proudfoot, N.J. (2000) Transcriptional termination and coupled polyadenylation in vitro. *EMBO J*, **19**, 3770-3777.

Yuryev, A. and Corden, J.L. (1996) Suppression analysis reveals a functional difference between the serines in positions two and five in the consensus sequence of the C- terminal domain of yeast RNA polymerase II. *Genetics*, **143**, 661-671.

Yuryev, A., Patturajan, M., Litingtung, Y., Joshi, R.V., Gentile, C., Gebara, M. and Corden, J.L. (1996) The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins. *Proc Natl Acad Sci U S A*, **93**, 6975-6980.

Zenklusen, D., Vinciguerra, P., Strahm, Y. and Stutz, F. (2001) The yeast hnRNP-Like proteins Yra1p and Yra2p participate in mRNA export through interaction with Mex67p. *Mol Cell Biol*, **21**, 4219-4232.

Zenklusen, D., Vinciguerra, P., Wyss, J.C. and Stutz, F. (2002) Stable mRNP formation and export require cotranscriptional recruitment of the mRNA export factors Yra1p and Sub2p by Hpr1p. *Mol Cell Biol*, **22**, 8241-8253.

Zhang, M. and Green, M.R. (2001) Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. *Genes Dev*, **15**, 30-35.

Zhang, Z., Fu, J. and Gilmour, D.S. (2005) CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes Dev*, **19**, 1572-1580.

Zhang, Z. and Gilmour, D.S. (2006) Pcf11 is a termination factor in Drosophila that dismantles the elongation complex by bridging the CTD of RNA polymerase II to the nascent transcript. *Mol Cell*, **21**, 65-74.

Zhang, Z., Klatt, A., Henderson, A.J. and Gilmour, D.S. (2007) Transcription termination factor Pcf11 limits the processivity of Pol II on an HIV provirus to repress gene expression. *Genes Dev*, **21**, 1609-1614.

Zhao, J., Hyman, L. and Moore, C. (1999) Formation of mRNA 3' ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol Mol Biol Rev*, **63**, 405-445.

Zhong, S., Salomoni, P. and Pandolfi, P.P. (2000) The transcriptional role of PML and the nuclear body. *Nat Cell Biol*, **2**, E85-90.

Zhou, Z., Luo, M.J., Straesser, K., Katahira, J., Hurt, E. and Reed, R. (2000) The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature*, **407**, 401-405.

Zorio, D.A. and Bentley, D.L. (2004) The link between mRNA processing and transcription: communication works both ways. *Exp Cell Res*, **296**, 91-97.